GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 23, 2005, 18:56:58 ; Search time 167 Seconds
(without alignments)
1185.756 Million cell updates/sec

Title: US-10-081-816-12
Perfect score: 2645
Sequence: 1 MRPSGEKVVKGHGQGNSGHS.....TTTVGTTLPPPIMKLKGRKG 512

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2645	100.0	512	4	ABB70017	Abb70017 Drosophil
2	2645	100.0	512	5	AAE30492	Aae30492 Fruit fly
3	405	15.3	454	4	ABB68814	Abb68814 Drosophil
4	390	14.7	364	4	AAB75197	Aab75197 Drosophil
5	390	14.7	364	5	AAE30504	Aae30504 Fruit fly
6	169	6.4	33	5	AAE30548	Aae30548 Fruit fly
7	160	6.0	433	4	AAB75232	Aab75232 Drosophil
8	135.5	5.1	369	5	AAE30509	Aae30509 Fruit fly
9	132.5	5.0	736	4	ABB67525	Abb67525 Drosophil
10	132.5	5.0	736	5	AAE30497	Aae30497 Fruit fly
11	132.5	5.0	1299	4	ABB70025	Abb70025 Drosophil
12	130.5	4.9	372	4	AAB75208	Aab75208 Drosophil
13	130.5	4.9	372	5	AAE30510	Aae30510 Fruit fly
14	129	4.9	436	4	ABB68743	Abb68743 Drosophil
15	129	4.9	436	5	AAE30491	Aae30491 Fruit fly
16	127	4.8	348	4	AAB75207	Aab75207 Drosophil
17	122	4.6	410	4	ABB71166	Abb71166 Drosophil
18	122	4.6	410	5	AAE30481	Aae30481 Fruit fly
19	122	4.6	414	4	AAB75193	Aab75193 Drosophil
20	120	4.5	450	5	AAE30537	Aae30537 Fruit fly
21	119	4.5	477	4	AAB75227	Aab75227 Drosophil
22	118.5	4.5	2008	4	AAU04090	Aau04090 Human cad
23	118.5	4.5	2008	5	ABG61703	Abg61703 Human cad
24	118.5	4.5	2008	6	ABR43636	Abr43636 Human CLA
25	118.5	4.5	2008	7	ADC79417	Adc79417 Human CLA
26	115	4.3	367	5	AAE30493	Aae30493 Fruit fly
27	115	4.3	817	4	ABB70027	Abb70027 Drosophil
28	114	4.3	693	4	ABB47271	Abb47271 Enterococ
29	113	4.3	250	4	ABB67792	Abb67792 Drosophil
30	113	4.3	378	7	ADD15295	Add15295 Fruitfly
31	112.5	4.3	365	5	AAE30536	Aae30536 Fruit fly
32	112.5	4.3	408	4	AAB75217	Aab75217 Drosophil
33	112.5	4.3	698	6	ABM67593	Abm67593 Photorhab
34	112.5	4.3	1227	4	ABB68608	Abb68608 Drosophil
35	110.5	4.2	408	5	AAE30517	Aae30517 Fruit fly
36	110.5	4.2	2073	6	ABR43629	Abr43629 Mouse CLA
37	110	4.2	498	4	ABB63071	Abb63071 Drosophil
38	110	4.2	498	5	AAE30495	Aae30495 Fruit fly
39	109.5	4.1	327	8	ADK16371	Adk16371 Nanoarcha
40	108	4.1	381	4	AAB75209	Aab75209 Drosophil
41	108	4.1	381	5	AAE30511	Aae30511 Fruit fly
42	108	4.1	408	5	ABP26074	Abp26074 Streptoco
43	107.5	4.1	346	6	ABU49652	Abu49652 Protein e
44	107	4.0	404	5	AAE30532	Aae30532 Fruit fly
45	106.5	4.0	390	4	AAB75221	Aab75221 Drosophil

ALIGNMENTS

RESULT 1
ABB70017
ID ABB70017 standard; protein; 512 AA.
XX
AC ABB70017;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 36843.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL14120.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 36843; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

```
SQ   Sequence 512 AA;

  Query Match             100.0%;  Score 2645;  DB 4;  Length 512;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-302;
  Matches  512;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRPSGEKVVKGHGQGNSGHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRPSGEKVVKGHGQGNSGHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSI 60

Qy         61 GLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGP 120

Qy        121 ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPI 180

Qy        181 MIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 MIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVI 240

Qy        241 THVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAM 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 THVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAM 300

Qy        301 VADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLT 360

Qy        361 ITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPS 420

Qy        421 TINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLD 480

Qy        481 NDMSLMGASTLSTTTVGTTLPPPIMKLKGRKG 512
              ||||||||||||||||||||||||||||||||
Db        481 NDMSLMGASTLSTTTVGTTLPPPIMKLKGRKG 512


RESULT 2
AAE30492
ID   AAE30492 standard; protein; 512 AA.
XX
AC   AAE30492;
XX
DT   24-FEB-2003  (first entry)
XX
DE   Fruit fly gustatory receptor protein, Gr63F1.
XX
KW   Fruit fly; gustatory receptor; Gr; odorant receptor; Or; pesticide;
KW   Gr63F1.
XX
OS   Drosophila melanogaster.
XX
PN   WO200268593-A2.
XX
PD   06-SEP-2002.
XX
PF   22-FEB-2002; 2002WO-US005414.
XX
PR   23-FEB-2001; 2001US-0271319P.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Axel R,  Scott K;
XX
DR   WPI; 2002-698668/75.
XX
PT   Novel nucleic acid encoding insect gustatory or odorant receptor protein
PT   useful for identifying a compound which specifically binds to the
PT   receptor for controlling a pest population in an area.
XX
PS   Claim 3; Page 160-162; 264pp; English.
XX
CC   The invention relates to an isolated nucleic acid encoding an insect
CC   gustatory receptor (Gr) or odorant receptor (Or) protein which comprises
CC   seven transmembrane domains and a C-terminal domain comprising
CC   consecutive amino acids. The invention is useful for identifying a
CC   compound which activates the insect receptor or inhibits the activity of
CC   the insect receptor. The purified insect receptor protein is embedded in
CC   a lipid bilayer. The invention is sprayed for combating ingestion of
CC   crops by pest insects, combating disease-carrying insects in an area and
CC   controlling a pest population in an area. The invention is useful for
CC   detecting the presence of insect gustatory or odorant receptor and for
CC   inhibiting the function of the receptor in humans or animals or in
CC   biological fluids isolated from them. The invention is also useful for
CC   identifying or isolating other insect receptors and for combating pest
CC   nuisances. The present sequence is fruit fly gustatory receptor (Gr)
CC   protein
XX
SQ   Sequence 512 AA;

  Query Match             100.0%;  Score 2645;  DB 5;  Length 512;
  Best Local Similarity   100.0%;  Pred. No. 7.7e-302;
  Matches  512;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRPSGEKVVKGHGQGNSGHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRPSGEKVVKGHGQGNSGHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSI 60

Qy         61 GLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGP 120

Qy        121 ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPI 180

Qy        181 MIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 MIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVI 240

Qy        241 THVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAM 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 THVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAM 300

Qy        301 VADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLT 360

Qy        361 ITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPS 420

Qy        421 TINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLD 480

Qy        481 NDMSLMGASTLSTTTVGTTLPPPIMKLKGRKG 512
              ||||||||||||||||||||||||||||||||
Db        481 NDMSLMGASTLSTTTVGTTLPPPIMKLKGRKG 512


RESULT 3
ABB68814
ID   ABB68814 standard; protein; 454 AA.
XX
AC   ABB68814;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 33234.
```

XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12917.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 33234; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 454 AA;

Query Match 15.3%; Score 405; DB 4; Length 454;
Best Local Similarity 26.1%; Pred. No. 8.7e-38;
Matches 115; Conservative 89; Mismatches 190; Indels 46; Gaps 11;

```
Qy         42 PLNSANAQAYLYGVRKY--SIGLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRN 99
Db         25 PMLNPNQRQFLEDEVRYREKLKLMARGDAMEEVYVRKQETVDDPLELD----KHDSFYQT 80

Qy        100 IDPINWFLRIIGVLPIVRHGP----ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRI- 154
Db         81 TKSLLVLFQIMGVMPIHRNPPEKNLPRTGYSWGSKQVMWAIFIYSCQTTIVVLVLRERVK 140

Qy        155 HIVRSLSGPFEEAVIAYLFLVNILPIMIIPIL-WYEARKIAKLFNDWDDFEVLYYQISGH 213
Db        141 KFVTSPDKRFDEAIYNVIFISLLFTNFLLPVASWRHGPQVAIFKNMWTNYQYKFFKTTGS 200

Qy        214 SLPLKLRQKAVYIAIVLPILSVLS---VVITHVTMSDLNINQVVPYCILDN--------- 261
Db        201 ------------PIVFPNLYPLTWSLCVFSWLLSIAINLSQ---YFLQPDFRLWYTFAY 244

Qy        262 --LTAMLGAW---WFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTR 316
Db        245 YPIIAMLNCFCSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQ 304

Qy        317 DTGNALCYTFVFMSLYLFFIITLSIYGLMSQ-LSEGFGIKDIGLTITALWNIGLLFYICD 375
Db        305 QLGRAYSNMYGMYCLVIFFTTIIATYGSISEIIDHGATYKEVGLFVIVFYCMGLLYIICN 364

Qy        376 EAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLF 435
Db        365 EAHYASRKVGLDFQTKLLNINLTAVDAATQKEVEMLLVAINKNPPIMNLDGYANINRELI 424
```

```
Qy        436 KGLLTTMVTYLVVLLQFQIS 455
Db        425 TTNISFMATYLVVLLQFKIT 444
```

RESULT 4
AAB75197
ID AAB75197 standard; protein; 364 AA.
XX
AC AAB75197;
XX
DT 03-APR-2001 (first entry)
XX
DE Drosophila gustatory receptor GR21D.1 protein sequence.
XX
KW Gustatory receptor; fruit fly; taste; pheromone; semiochemical;
KW crop damage; pest control.
XX
OS Drosophila melanogaster.
XX
PN WO200077208-A2.
XX
PD 21-DEC-2000.
XX
PF 14-JUN-2000; 2000WO-US016211.
XX
PR 14-JUN-1999; 99US-0138668P.
PR 10-FEB-2000; 2000US-0181704P.
XX
PA (UYYA) UNIV YALE.
XX
PI Carlson PJ, Clyne PJ, Warr CG;
XX
DR WPI; 2001-061873/07.
DR N-PSDB; AAF63736.
XX
PT New isolated nucleic acid molecule encoding Drosophila Gustatory Receptor
PT protein useful for e.g. identification of compounds which may be used for
PT pest management.
XX
PS Claim 12; Page 90-91; 227pp; English.
XX
CC This invention relates to polynucleotide sequences AAF63732 - AAF63777
CC which encode Drosophila gustatory receptor proteins represented by
CC sequences AAB75193 - AAB75238. The invention includes methods for
CC determining gustatory receptor ligands. Also included is a method for
CC modulating the expression of the DNA encoding the receptors. The DNA and
CC protein sequences may be used for the identification of compounds, e.g.
CC pheromones and other semiochemicals, which may be used for pest
CC management. The DNA sequences may also be used for behavioural studies
CC involving gustatory systems in various organisms. Also, the DNA sequences
CC may also be used to track down gustatory receptor genes in insects that
CC damage crops or transmit diseases
XX
SQ Sequence 364 AA;

Query Match 14.7%; Score 390; DB 4; Length 364;
Best Local Similarity 27.3%; Pred. No. 3.6e-36;
Matches 101; Conservative 77; Mismatches 152; Indels 40; Gaps 9;

```
Qy        110 IGVLPIVRHGP----ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRI-HIVRSLSGPF 164
Db          1 MGVMPIHRNPPEKNLPRTGYSWGSKQVMWAIFIYSCQTTIVVLVLRERVKKFVTSPDKRF 60

Qy        165 EEAVIAYLFLVNILPIMIIPIL-WYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKA 223
Db         61 DEAIYNVIFISLLFTNFLLPVASWRHGPQVAIFKNMWTNYQYKFFKTTGS---------- 110

Qy        224 VYIAIVLPILSVLS---VVITHVTMSDLNINQVVPYCILDN----------LTAMLGAW 269
Db        111 ---PIVFPNLYPLTWSLCVFSWLLSIAINLSQ---YFLQPDFRLWYTFAYYPIIAMLNCF 164

Qy        270 ---WFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTF 326
```

```
                |:: | |    :  |::  |  ::   ||  : :|| ||: || : :  | |    :
Db        165 CSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMY 224

Qy        327 VFMSLYLFFIITLSIYGLMSQ-LSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVR 385
              | :||   :: || :|: :  |   |::|| :   : :|||: ||:||||||  |
Db        225 GMYCLVIFFTTIIATYGSISEIIDHGATYKEVGLFVIVFYCMGLLYIICNEAHYASRKVG 284

Qy        386 TNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTY 445
             :|| ||| : |  :::  | |: | | |   || :|  |: ::|| |    :: | ||
Db        285 LDFQTKLLNINLTAVDAATQKEVEMLLVAINKNPPIMNLDGYANINRELITTNISFMATY 344

Qy        446 LVVLLQFQIS 455
              |||||||:|:
Db        345 LVVLLQFKIT 354


RESULT 5
AAE30504
ID   AAE30504 standard; protein; 364 AA.
XX
AC   AAE30504;
XX
DT   24-FEB-2003  (first entry)
XX
DE   Fruit fly gustatory receptor protein, Gr21D1.
XX
KW   Fruit fly; gustatory receptor; Gr; odorant receptor; Or; pesticide;
KW   Gr21D1.
XX
OS   Drosophila melanogaster.
XX
PN   WO200268593-A2.
XX
PD   06-SEP-2002.
XX
PF   22-FEB-2002; 2002WO-US005414.
XX
PR   23-FEB-2001; 2001US-0271319P.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Axel R,  Scott K;
XX
DR   WPI; 2002-698668/75.
XX
PT   Novel nucleic acid encoding insect gustatory or odorant receptor protein
PT   useful for identifying a compound which specifically binds to the
PT   receptor for controlling a pest population in an area.
XX
PS   Disclosure; Page 183-185; 264pp; English.
XX
CC   The invention relates to an isolated nucleic acid encoding an insect
CC   gustatory receptor (Gr) or odorant receptor (Or) protein which comprises
CC   seven transmembrane domains and a C-terminal domain comprising
CC   consecutive amino acids. The invention is useful for identifying a
CC   compound which activates the insect receptor or inhibits the activity of
CC   the insect receptor. The purified insect receptor protein is embedded in
CC   a lipid bilayer. The invention is sprayed for combating ingestion of
CC   crops by pest insects, combating disease-carrying insects in an area and
CC   controlling a pest population in an area. The invention is useful for
CC   detecting the presence of insect gustatory or odorant receptor and for
CC   inhibiting the function of the receptor in humans or animals or in
CC   biological fluids isolated from them. The invention is also useful for
CC   identifying or isolating other insect receptors and for combating pest
CC   nuisances. The present sequence is fruit fly gustatory receptor (Gr)
CC   protein
XX
SQ   Sequence 364 AA;

  Query Match             14.7%;  Score 390;  DB 5;  Length 364;
  Best Local Similarity   27.3%;  Pred. No. 3.6e-36;
  Matches  101;  Conservative   77;  Mismatches  152;  Indels   40;  Gaps    9;

Qy        110 IGVLPIVRHGP----ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRI-HIVRSLSGPF 164
              :||:|| |: |      |  :   |   ::::  :       |  |   |:   | |     |
Db          1 MGVMPIHRNPPEKNLPRTGYSWGSKQVMWAIFIYSCQTTIVVLVLRERVKKFVTSPDKRF 60

Qy        165 EEAVIAYLFLVNILPIMIIPIL-WYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKA 223
              :||:   :|:  :    ::|:  |    ::|   | | :::  ::: :|
Db         61 DEAIYNVIFISLLFTNFLLPVASWRHGPQVAIFKNMWTNYQYKFFKTTGS---------- 110

Qy        224 VYIAIVLPILSVLS---VVITHVTMSDLNINQVVPYCILDN----------LTAMLGAW 269
                || | |  :    | : :   :|::|   | :  :           : |||  :
Db        111 ---PIVFPNLYPLTWSLCVFSWLLSIAINLSQ---YFLQPDFRLWYTFAYYPIIAMLNCF 164

Qy        270 ---WPLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTF 326
                 |:: | |    :  |::  |  ::   ||  : :|| ||: || : :  | |    :
Db        165 CSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMY 224

Qy        327 VFMSLYLFFIITLSIYGLMSQ-LSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVR 385
              | :||   :: || :|: :  |   |::|| :   : :|||: ||:||||||  |
Db        225 GMYCLVIFFTTIIATYGSISEIIDHGATYKEVGLFVIVFYCMGLLYIICNEAHYASRKVG 284

Qy        386 TNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTY 445
             :|| ||| : |  :::  | |: | | |   ||  :|  |: ::|| |    :: | ||
Db        285 LDFQTKLLNINLTAVDAATQKEVEMLLVAINKNPPIMNLDGYANINRELITTNISFMATY 344

Qy        446 LVVLLQFQIS 455
              |||||||:|:
Db        345 LVVLLQFKIT 354


RESULT 6
AAE30548
ID   AAE30548 standard; peptide; 33 AA.
XX
AC   AAE30548;
XX
DT   24-FEB-2003  (first entry)
XX
DE   Fruit fly gustatory receptor peptide, Gr63F1.
XX
KW   Fruit fly; gustatory receptor; Gr; odorant receptor; Or; pesticide;
KW   Gr63F1.
XX
OS   Drosophila melanogaster.
XX
PN   WO200268593-A2.
XX
PD   06-SEP-2002.
XX
PF   22-FEB-2002; 2002WO-US005414.
XX
PR   23-FEB-2001; 2001US-0271319P.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Axel R,  Scott K;
XX
DR   WPI; 2002-698668/75.
XX
PT   Novel nucleic acid encoding insect gustatory or odorant receptor protein
PT   useful for identifying a compound which specifically binds to the
PT   receptor for controlling a pest population in an area.
XX
PS   Disclosure; Page 249; 264pp; English.
XX
CC   The invention relates to an isolated nucleic acid encoding an insect
CC   gustatory receptor (Gr) or odorant receptor (Or) protein which comprises
CC   seven transmembrane domains and a C-terminal domain comprising
CC   consecutive amino acids. The invention is useful for identifying a
CC   compound which activates the insect receptor or inhibits the activity of
CC   the insect receptor. The purified insect receptor protein is embedded in
CC   a lipid bilayer. The invention is sprayed for combating ingestion of
CC   crops by pest insects, combating disease-carrying insects in an area and
CC   controlling a pest population in an area. The invention is useful for
```

```
CC   detecting the presence of insect gustatory or odorant receptor and for
CC   inhibiting the function of the receptor in humans or animals or in
CC   biological fluids isolated from them. The invention is also useful for
CC   identifying or isolating other insect receptors and for combating pest
CC   nuisances. The present sequence is fruit fly gustatory receptor (Gr)
CC   peptide
XX
SQ   Sequence 33 AA;

  Query Match             6.4%;  Score 169;  DB 5;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 8e-12;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        420 STINCGGFFDVNRTLFKGLLTTMVTYLVVLLQF 452
              |||||||||||||||||||||||||||||||||
Db          1 STINCGGFFDVNRTLFKGLLTTMVTYLVVLLQF 33


RESULT 7
AAB75232
ID   AAB75232 standard; protein; 433 AA.
XX
AC   AAB75232;
XX
DT   03-APR-2001  (first entry)
XX
DE   Drosophila gustatory receptor GR98B.1 protein sequence.
XX
KW   Gustatory receptor; fruit fly; taste; pheromone; semiochemical;
KW   crop damage; pest control.
XX
OS   Drosophila melanogaster.
XX
PN   WO200077208-A2.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-US016211.
XX
PR   14-JUN-1999;   99US-0138668P.
PR   10-FEB-2000; 2000US-0181704P.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Carlson PJ,  Clyne PJ,  Warr CG;
XX
DR   WPI; 2001-061873/07.
DR   N-PSDB; AAF63771.
XX
PT   New isolated nucleic acid molecule encoding Drosophila Gustatory Receptor
PT   protein useful for e.g. identification of compounds which may be used for
PT   pest management.
XX
PS   Claim 12; Page 214-215; 227pp; English.
XX
CC   This invention relates to polynucleotide sequences AAF63732 - AAF63777
CC   which encode Drosophila gustatory receptor proteins represented by
CC   sequences AAB75193 - AAB75238. The invention includes methods for
CC   determining gustatory receptor ligands. Also included is a method for
CC   modulating the expression of the DNA encoding the receptors. The DNA and
CC   protein sequences may be used for the identification of compounds, e.g.
CC   pheromones and other semiochemicals, which may be used for pest
CC   management. The DNA sequences may also be used for behavioural studies
CC   involving gustatory systems in various organisms. Also, the DNA sequences
CC   may also be used to track down gustatory receptor genes in insects that
CC   damage crops or transmit diseases
XX
SQ   Sequence 433 AA;

  Query Match             6.0%;  Score 160;  DB 4;  Length 433;
  Best Local Similarity   20.9%;  Pred. No. 6.3e-09;
  Matches   91;  Conservative   72;  Mismatches  139;  Indels  134;  Gaps   18;

Qy        102 PINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNR-IHIVRS- 159
              || :| | :       |   |    :  | :: |:   |:  ||   ||  : ||   :
Db         27 PIQFFTRTL-------HKRRRGIVILGYACYLISISLMVIYECYANIVALQKDIHKFHAE 79

Qy        160 ----LSGPF-EEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEV-LYYQISG- 212
                  : | : | :    :|: | | |::         |::|::::|  | |: |    ||
Db         80 DSSKVMGEYAERSWWVAMFVWNQLNILL------NFRRLARIYDDIADLEIDLNNASSGF 133

Qy        213 ----HSLPLKLR-QKAVYIAIVL-------------PILSVLSVVITHVTMSDLNINQV 253
                  |    : |   :| : |||              | |   : |:| : :  | : :
Db        134 VGQRHWWRFRFRLALSVGLWIVLLVGLTPRFTLVALGPYLHWTNKVLTEIILIMLQL-KC 192

Qy        254 VPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSK 313
                ||:          :  || | :    |:| ::    |:
Db        193 TEYCV----------FVLLIYELILRGRHIL-QQISVELE-----------------GN 223

Qy        314 LTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYI 373
              :||:   ||     ::|   ::   |:||::::|  |    : ||:  |:|   :  |
Db        224 QSRDSVQELC-----VALKRNQLLAGRIWGLVNEVSLYF---TLSLTLLFLYNELTILQI 275

Qy        374 CDEAHYASVN-------------VRTNFQKKLLMVELNWMNSDAQTEINMFLRA--TEMN 418
              : |   |||               : ||  | ::  :  : :     ||:||    :|
Db        276 VNWALIKSVNPNECCQYSKLVFKFKRNFTYKQVIFIIGRVGTCLLLSINIFLSCLYSEFC 335

Qy        419 PSTIN----------------------------------------CGGFFDVNRTLFKGL 438
              | |                                        ||| ||:|   | |:
Db        336 IQTYNSISRVLHQMYCLSAAEDYLILKMGLREYSLQMEHLKLIFTCGGLFDINLKFFGGM 395

Qy        439 LTTMVTYLVVLLQFQI 454
              : |:  |:::|:||:|
Db        396 VVTLFGYIIILVQFKI 411


RESULT 8
AAE30509
ID   AAE30509 standard; protein; 369 AA.
XX
AC   AAE30509;
XX
DT   24-FEB-2003  (first entry)
XX
DE   Fruit fly gustatory receptor protein, Gr39D1.
XX
KW   Fruit fly; gustatory receptor; Gr; odorant receptor; Or; pesticide;
KW   Gr39D1.
XX
OS   Drosophila melanogaster.
XX
PN   WO200268593-A2.
XX
PD   06-SEP-2002.
XX
PF   22-FEB-2002; 2002WO-US005414.
XX
PR   23-FEB-2001; 2001US-0271319P.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Axel R,  Scott K;
XX
DR   WPI; 2002-698668/75.
XX
PT   Novel nucleic acid encoding insect gustatory or odorant receptor protein
PT   useful for identifying a compound which specifically binds to the
PT   receptor for controlling a pest population in an area.
XX
PS   Disclosure; Page 193-194; 264pp; English.
XX
CC   The invention relates to an isolated nucleic acid encoding an insect
CC   gustatory receptor (Gr) or odorant receptor (Or) protein which comprises
CC   seven transmembrane domains and a C-terminal domain comprising
CC   consecutive amino acids. The invention is useful for identifying a
CC   compound which activates the insect receptor or inhibits the activity of
```

CC the insect receptor. The purified insect receptor protein is embedded in
CC a lipid bilayer. The invention is sprayed for combating ingestion of
CC crops by pest insects, combating disease-carrying insects in an area and
CC controlling a pest population in an area. The invention is useful for
CC detecting the presence of insect gustatory or odorant receptor and for
CC inhibiting the function of the receptor in humans or animals or in
CC biological fluids isolated from them. The invention is also useful for
CC identifying or isolating other insect receptors and for combating pest
CC nuisances. The present sequence is fruit fly gustatory receptor (Gr)
CC protein
XX
SQ Sequence 369 AA;

Query Match 5.1%; Score 135.5; DB 5; Length 369;
Best Local Similarity 20.5%; Pred. No. 3.7e-06;
Matches 87; Conservative 65; Mismatches 126; Indels 147; Gaps 20;

```
Qy      106 FLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFE 165
Db       11 YFALLGLVP-WSESCAQSKFVQK----VYSAILIIL----------NAVHFGISIYFP-Q 54

Qy      166 EAVIAYLFLVNIL----PIMIIPILWYEARKIAKLFNDWDDF-----EVLYY------QI 210
Db       55 SAELFLSLMVNVIVFVARIVCVTVI------ILQVMVHYDDYFRFCREMKYLGLRLQCEL 108

Qy      211 SGHSLPLKLRQKAVYIAI-------VLPIL-------------SVLSVVIT--------- 241
Db      109 KIHVGRLKWQSYAKILALGIGFLVTVLPSIYVALSGSLLYFWSSLLSILIIRMQFVLVLL 168

Qy      242 -------HVTMSDLNINQVV------PYCILDNLTAMLGAWWFLICEAMS-ITAHLLAER 287
Db      169 NVELLGHHVSLLGIRLQNVLECHLMGANCTLDGNANRLCSLEFLLALKQSHMQLHYLFTH 228

Qy      288 FQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQ 347
Db      229 FNDLF-------------------------GWSILGTYV----VLFSDSTVNIYWTQQV 258

Qy      348 LSEGFGIKDIGLT----ITALWNIGLLFYICDE----------------AHYASVNVRTN 387
Db      259 LVEVYEYKYLYATFSVFVPSFFNI-LVFCRCGEFCQRQSVLIGSYLRNLSCHPSIGRETS 317

Qy      388 FQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLV 447
Db      318 Y-KDLLM--------------EFILQVEQNVLAINAEGFMSTDNSLLMSILAAKVTYLI 361

Qy      448 VLLQF 452
Db      362 VLMQF 366
```

RESULT 9
ABB67525
ID ABB67525 standard; protein; 736 AA.
XX
AC ABB67525;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 29367.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL11628.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 29367; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 736 AA;

Query Match 5.0%; Score 132.5; DB 4; Length 736;
Best Local Similarity 21.0%; Pred. No. 2.6e-05;
Matches 77; Conservative 66; Mismatches 134; Indels 89; Gaps 16;

```
Qy      102 PINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNR-IHIVRS- 159
Db      427 PIQFFTRTL-------HKRRRGIVILGYACYLISISLMVIYECYANIVALQKDIHKFHAE 479

Qy      160 ----LSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEV-LYYQISG-- 212
Db      480 DSSKVMGNTQKVLVVAMFVWNQLNILL------NFRRLARIYDDIADLEIDLNNASSGFV 533

Qy      213 ---HSLPLKLR-QKAVYIAIVL--------------PILSVLSVVITHVTMSDLNINQVV 254
Db      534 GQRHWWRFRFRLALSVGLWIVLLVGLTPRFTLVALGPYLHWTNKVLTEIILIMLQL-KCT 592

Qy      255 PYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKL 314
Db      593 EYCV----------FVLLIYELILRGRHIL-QQISVELE-----------------GNQ 623

Qy      315 TRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYIC 374
Db      624 SRDSVQELC-----VALKRNQLLAGRIWGLVNEVSLYF---TLSLTLLFLYNELTILQIV 675

Qy      375 DEAHYASVNVR--TNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNR 432
Db      676 NWALIKSVNPNECCQYTEDYLILKMGLREYSLQMEHLKLI---------FTCGGLFDINL 726

Qy      433 TLFKGL 438
Db      727 KFFGGV 732
```

RESULT 10
AAE30497
ID AAE30497 standard; protein; 736 AA.
XX
AC AAE30497;
XX
DT 24-FEB-2003 (first entry)
XX
DE Fruit fly gustatory receptor protein, Gr98A1.
XX
KW Fruit fly; gustatory receptor; Gr; odorant receptor; Or; pesticide;
KW Gr98A1.
XX

OS Drosophila melanogaster.
XX
PN WO200268593-A2.
XX
PD 06-SEP-2002.
XX
PF 22-FEB-2002; 2002WO-US005414.
XX
PR 23-FEB-2001; 2001US-0271319P.
XX
PA (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI Axel R, Scott K;
XX
DR WPI; 2002-698668/75.
XX
PT Novel nucleic acid encoding insect gustatory or odorant receptor protein
PT useful for identifying a compound which specifically binds to the
PT receptor for controlling a pest population in an area.
XX
PS Claim 3; Page 170-173; 264pp; English.
XX
CC The invention relates to an isolated nucleic acid encoding an insect
CC gustatory receptor (Gr) or odorant receptor (Or) protein which comprises
CC seven transmembrane domains and a C-terminal domain comprising
CC consecutive amino acids. The invention is useful for identifying a
CC compound which activates the insect receptor or inhibits the activity of
CC the insect receptor. The purified insect receptor protein is embedded in
CC a lipid bilayer. The invention is sprayed for combating ingestion of
CC crops by pest insects, combating disease-carrying insects in an area and
CC controlling a pest population in an area. The invention is useful for
CC detecting the presence of insect gustatory or odorant receptor and for
CC inhibiting the function of the receptor in humans or animals or in
CC biological fluids isolated from them. The invention is also useful for
CC identifying or isolating other insect receptors and for combating pest
CC nuisances. The present sequence is fruit fly gustatory receptor (Gr)
CC protein
XX
SQ Sequence 736 AA;

Query Match 5.0%; Score 132.5; DB 5; Length 736;
Best Local Similarity 21.0%; Pred. No. 2.6e-05;
Matches 77; Conservative 66; Mismatches 134; Indels 89; Gaps 16;

```
Qy        102 PINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNR-IHIVRS- 159
              || :| | :       |   |    :  | :: |:   |:  ||   ||  : ||   :
Db        427 PIQFFTRTL-------HKRRRGIVILGYACYLISISLMVIYECYANIVALQKDIHKFHAE 479

Qy        160 ----LSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEV-LYYQISG-- 212
                  : |  :: ::  :|: | | |::       |::|::::|  | |: |    ||
Db        480 DSSKVMGNTQKVLVVAMFVWNQLNILL------NFRRLARIYDDIADLEIDLNNASSGFV 533

Qy        213 ---HSLPLKLR-QKAVYIAIVL-------------PILSVLSVVITHVTMSDLNINQVV 254
                  |    : |   :| : |||              | |   : |:| : :  | : :
Db        534 GQRHWWRFRFRLALSVGLWIVLLVGLTPRFTLVALGPYLHWTNKVLTEIILIMLQL-KCT 592

Qy        255 PYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKL 314
              ||:          :  || | :    |:| ::    |:
Db        593 EYCV----------FVLLIYELILRGRHIL-QQISVELE-----------------GNQ 623

Qy        315 TRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYIC 374
              :||:   ||     ::|   ::   |:||::::|  |    : ||:  |:|   :  |
Db        624 SRDSVQELC-----VALKRNQLLAGRIWGLVNEVSLYF---TLSLTLLFLYNELTILQIV 675

Qy        375 DEAHYASVNVR--TNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNR 432
              : |   |||      : :  |::::       | |    :          ||| ||:|
Db        676 NWALIKSVNPNECCQYTEDYLILKMGLREYSLQMEHLKLI---------FTCGGLFDINL 726

Qy        433 TLFKGL 438
                | |:
Db        727 KFFGGV 732
```

RESULT 11
ABB70025
ID ABB70025 standard; protein; 1299 AA.
XX
AC ABB70025;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 36867.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL14128.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 36867; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1299 AA;

Query Match 5.0%; Score 132.5; DB 4; Length 1299;
Best Local Similarity 19.3%; Pred. No. 6.6e-05;
Matches 81; Conservative 75; Mismatches 159; Indels 105; Gaps 19;

```
Qy         96 FYRNIDPINWFLRIIGVLPI--VRH-GPARAKFEMNSASFIYSVVFFVLLACYVGYVANN 152
              |:| :  : :  :| |:||:  ||     |  :|    |   :||::  :|     | | |
Db        484 FHRAVSNVLFISQIYGLLPVSNVRALDVADIRFRWCSPRILYSLLIGILNLSEPGAVINY 543

Qy        153 RIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEVLYYQISG 212
              | :  :    |  :    |::| :|  :     |  | :  ::   |   | |: ::
Db        544 VIKVTIN----FHTSSTLSLYIVCLLEHL---FFWRLAIQWPRIMRTWHGVEQLFLRVPY 596

Qy        213 HSLPLKLRQKAVYIAIVLPILSVL---------SVVITHVTMSDLNIN------------ 251
                    :: :||   : : | |           |  ::::  :   ||
Db        597 RFYGEYRIKRRIYIVFTIVMSSALVEHCLLLGNSFHLSNMERTQCKINVTYFESIYKWER 656

Qy        252 ----QVVPYCILDNLTAMLGAW------W--------------FLICEAMSITA--HLLA 285
                  ::||             |      |              |::|  : : |  | |
Db        657 PHLYMILPYHF----------WMLPILEWVNQTIAYPRSFTDCFIMCIGIGLAARFHQLY 706

Qy        286 ERFQKALKHIGPAAMVADYRVLWLRLSKLTR--DTGNALCYTFVF---MSLYLFFIITLS 340
              |     : : ||   : |  :| | :|   |  |      |   ||   |
Db        707 RRIAAVHRKVMPAVFWTEVREHYLALKRLVHLLDAAIAPLVLLAFGNNMSFICF------ 760
```

```
Qy        341 IYGLMSQLSEGFGIKDIG---LTITALW-NIG-------LLFYICDEAHYASVNVRTNFQ 389
                ||   |  |:||   | : | | ::|       |  ::     :|:|   :::
Db        761 ------QLFNSF--KNIGVDFLVMLAFWYSLGFAVVRTLLTIFVA-----SSIN---DYE 804

Qy        390 KKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVL 449
              :|::    :  :    |:  |     : : ::  ||| : |:|   : ||::|| :::
Db        805 RKIVTALRDVPSRAWSIEVQRFSEQLGNDTTALSGSGFFYLTRSLVLAMGTTIITYELMI 864


RESULT 12
AAB75208
ID   AAB75208 standard; protein; 372 AA.
XX
AC   AAB75208;
XX
DT   03-APR-2001  (first entry)
XX
DE   Drosophila gustatory receptor GR39D.2a protein sequence.
XX
KW   Gustatory receptor; fruit fly; taste; pheromone; semiochemical;
KW   crop damage; pest control.
XX
OS   Drosophila melanogaster.
XX
PN   WO200077208-A2.
XX
PD   21-DEC-2000.
XX
PF   14-JUN-2000; 2000WO-US016211.
XX
PR   14-JUN-1999;   99US-0138668P.
PR   10-FEB-2000; 2000US-0181704P.
XX
PA   (UYYA ) UNIV YALE.
XX
PI   Carlson PJ,  Clyne PJ,  Warr CG;
XX
DR   WPI; 2001-061873/07.
DR   N-PSDB; AAF63747.
XX
PT   New isolated nucleic acid molecule encoding Drosophila Gustatory Receptor
PT   protein useful for e.g. identification of compounds which may be used for
PT   pest management.
XX
PS   Claim 12; Page 129-130; 227pp; English.
XX
CC   This invention relates to polynucleotide sequences AAF63732 - AAF63777
CC   which encode Drosophila gustatory receptor proteins represented by
CC   sequences AAB75193 - AAB75238. The invention includes methods for
CC   determining gustatory receptor ligands. Also included is a method for
CC   modulating the expression of the DNA encoding the receptors. The DNA and
CC   protein sequences may be used for the identification of compounds, e.g.
CC   pheromones and other semiochemicals, which may be used for pest
CC   management. The DNA sequences may also be used for behavioural studies
CC   involving gustatory systems in various organisms. Also, the DNA sequences
CC   may also be used to track down gustatory receptor genes in insects that
CC   damage crops or transmit diseases
XX
SQ   Sequence 372 AA;

  Query Match             4.9%;  Score 130.5;  DB 4;  Length 372;
  Best Local Similarity   21.7%;  Pred. No. 1.5e-05;
  Matches   80;  Conservative   64;  Mismatches  135;  Indels   89;  Gaps   17;

Qy        129 SASFIYSVVFFVLLACYVGYVANNRIHI-VRSLSG-PFEEAVI---------AYLFL--- 174
              |::  |  ||  | :|  :| :| :  ::  :| :|  |: |||         | |:|
Db         39 SSTAIQIVVVGVFMAALLGALAESLYYMETKSQTGNTFDNAVILTTSVTQLLANLWLRSQ 98

Qy        175 ----VNILPIM--IIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAI 228
                  ||:|  :  :: :| :|   : :       |  ||                   |
Db         99 QKSQVNLLQRLSQVVELLQFEPYAVPQ-------FRWLY------------------RI 132

Qy        229 VLPILSVLSVVITHVTMSDLNINQV---------VPYCILDNLTAMLGAWWFLICEAMSI 279
               | :  :  ::||  :: |   |:          |  |:| |       : |     | :
Db        133 WLLVCLIYGAMVTHFGINWLTTMQISRVLTLIGFVYRCVLAN-------FQFTCYTGMVV 185

Qy        280 TAHLLAERFQKALKHI---------GPAAMVADY-RVLWLRLSKLTRDTGNALCYTFVF- 328
                | :  | |:|:          | |  :  :  :| |   :|    |  : : |::
Db        186 ILKKLLQVQVKQLEHLVSTTTISMAGVAGCLRTHDEILLLGQRELIAVYGGVILFLFIYQ 245

Qy        329 --MSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVRT 386
                 : :|:|  |          |||   :  | :   |  :|||:       :: :
Db        246 VMQCILIFYISNL----------EGFHSSN-DLVLIFCWLAPMLFYLILPLVVNDIHNQA 294

Qy        387 NFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNR-TLFKGLLTTMVTY 445
              |   |:|  ::   :     |  ||         :   ||| ::: |||| | | : ||
Db        295 NKTAKML-TKVPRTGTGLDRMIEKFLLKNLRQKPILTAYGFFALDKSTLFK-LFTAIFTY 352

Qy        446 LVVLLQFQ 453
              :|:|:||:
Db        353 MVILVQFK 360


RESULT 13
AAE30510
ID   AAE30510 standard; protein; 372 AA.
XX
AC   AAE30510;
XX
DT   24-FEB-2003  (first entry)
XX
DE   Fruit fly gustatory receptor protein, Gr39D2a.
XX
KW   Fruit fly; gustatory receptor; Gr; odorant receptor; Or; pesticide;
KW   Gr39D2a.
XX
OS   Drosophila melanogaster.
XX
PN   WO200268593-A2.
XX
PD   06-SEP-2002.
XX
PF   22-FEB-2002; 2002WO-US005414.
XX
PR   23-FEB-2001; 2001US-0271319P.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
PI   Axel R,  Scott K;
XX
DR   WPI; 2002-698668/75.
XX
PT   Novel nucleic acid encoding insect gustatory or odorant receptor protein
PT   useful for identifying a compound which specifically binds to the
PT   receptor for controlling a pest population in an area.
XX
PS   Disclosure; Page 194-196; 264pp; English.
XX
CC   The invention relates to an isolated nucleic acid encoding an insect
CC   gustatory receptor (Gr) or odorant receptor (Or) protein which comprises
CC   seven transmembrane domains and a C-terminal domain comprising
CC   consecutive amino acids. The invention is useful for identifying a
CC   compound which activates the insect receptor or inhibits the activity of
CC   the insect receptor. The purified insect receptor protein is embedded in
CC   a lipid bilayer. The invention is sprayed for combating ingestion of
CC   crops by pest insects, combating disease-carrying insects in an area and
CC   controlling a pest population in an area. The invention is useful for
CC   detecting the presence of insect gustatory or odorant receptor and for
CC   inhibiting the function of the receptor in humans or animals or in
CC   biological fluids isolated from them. The invention is also useful for
CC   identifying or isolating other insect receptors and for combating pest
CC   nuisances. The present sequence is fruit fly gustatory receptor (Gr)
CC   protein
XX
SQ   Sequence 372 AA;
```

```
  Query Match             4.9%;  Score 130.5;  DB 5;  Length 372;
  Best Local Similarity   21.7%;  Pred. No. 1.5e-05;
  Matches   80;  Conservative   64;  Mismatches  135;  Indels   89;  Gaps   17;

Qy        129 SASPIYSVVFFVLLACYVGYVANNRIHI-VRSLSG-PFEEAVI---------AYLFL--- 174
              |::  |  ||  | :|  :| :| :  ::  :| :|  |: |||         | |:|
Db         39 SSTAIQIVVVGVFMAALLGALAESLYYMETKSQTGNTFDNAVILTTSVTQLLANLWLRSQ 98

Qy        175 ----VNILPIM--IIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAI 228
                  ||:|  :  :: :| :|   : :       |  ||                   |
Db         99 QKSQVNLLQRLSQVVELLQFEPYAVPQ-------FRWLY------------------RI 132

Qy        229 VLPILSVLSVVITHVTMSDLNINQV---------VPYCILDNLTAMLGAWWFLICEAMSI 279
               | :  :   ::||  :: |   |:         |  |:| |       : |      | :
Db        133 WLLVCLIYGAMVTHFGINWLTTMQISRVLTLIGFVYRCVLAN-------FQFTCYTGMVV 185

Qy        280 TAHLLAERFQKALKHI---------GPAAMVADY-RVLWLRLSKLTRDTGNALCYTFVF- 328
               | :   | |:|:          | |  :  :  :| |   :|     |  : : |::
Db        186 ILKKLLQVQVKQLEHLVSTTTISMAGVAGCLRTHDEILLLGQRELIAVYGGVILFLFIYQ 245

Qy        329 --MSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVRT 386
                 : :|:|  |           |||   :  | :   |   :|||:         :: :
Db        246 VMQCILIFYISNL----------EGFHSSN-DLVLIFCWLAPMLFYLILPLVVNDIHNQA 294

Qy        387 NFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNR-TLFKGLLTTMVTY 445
               |   |:|  ::    :     |  ||          :  ||| ::: |||| | | : ||
Db        295 NKTAKML-TKVPRTGTGLDRMIEKFLLKNLRQKPILTAYGFFALDKSTLFK-LFTAIFTY 352

Qy        446 LVVLLQFQ 453
               :|:|:||:
Db        353 MVILVQFK 360


RESULT 14
ABB68743
ID   ABB68743 standard; protein; 436 AA.
XX
AC   ABB68743;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 33021.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL12846.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 33021; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 436 AA;

  Query Match             4.9%;  Score 129;  DB 4;  Length 436;
  Best Local Similarity   21.2%;  Pred. No. 2.9e-05;
  Matches   94;  Conservative   70;  Mismatches  172;  Indels  108;  Gaps   22;

Qy         85 ASNNPEFKP-SVFYRNIDPINWFLRIIGVLP---IVRHGPARAKFEMNSASPIYSVVFFV 140
              |     ||:    ||  | |      :  |::|   |     |    ||:: |     : :| :
Db         30 AQGGLEFEQLDTFYGAIRPYLCVAQFFGIMPLSNIRSRDPQDVKFKVRSIGLAVTGLFLL 89

Qy        141 L--LACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYE----ARKIA 194
              |  :   ||  ||      :    |  : ::  :||:       :  : |      ||  :
Db         90 LGGMKTLVG--AN-----ILFTEGLNAKNIVGLVFLI------VGMVNWLNFVGFARSWS 136

Qy        195 KLFNDWDDFEVLY----YQISGHSLPLKLRQKAVYIAIVLPILSV--------------- 235
               :   |   ::|     |:    |    || |   :|: : :|:|
Db        137 HIMLPWSSVDILMLFPPYKRGKRS----LRSKVNVLALSVVVLAVGDHMLYYASGYCSYS 192

Qy        236 LSVVITHVTMSDL----------NINQVVPYCILDNLTAM-LGAW--------WFLICE 275
              : ::   |   | :          :|  ::|:    |: :|  | |         |  :
Db        193 MHILQCHTNHSRITFGLYLEKEFSDIMFIMPF----NIFSMCYGFWLNGAFTFLWNFMDI 248

Qy        276 AMSITAHLLAERFQKALKHIG-------PAAMVADYRVLWLRLSKLTRDTGNALCYTFVF 328
               : :|:  ||:|||:    :|       | |:  | |   :|| :|
Db        249 FIVMTSIGLAQRFQQFAARVGALEGRHVPEALWYDIRRDHIRLCELA-------SLVEAS 301

Qy        329 MSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALW--------NIGLLFYICDEAHYA 380
              ||  :|      ::| : :|     |       :      |              |:|    : | |
Db        302 MSNIVFVSCANNVYVICNQALAIFTKLRHPINYVYFWYSLIFLLARTSLVFMTASKIHDA 361

Qy        381 SVNVRTNFQKKLLMVELNWMNSDAQT-EINMFLRATEMNPSTINCGGF--FDVNRTLFKG 437
              |:          | :  |  : ||  | |:  |  | ::    :    |:  | : |    |
Db        362 SL---------LPLRSLYLVPSDGWTQEVQRF--ADQLTSEFVGLSGYRLFCLTRKSLFG 410

Qy        438 LLTTMVTYLVVLLQFQISIPTDKG 461
              :| |:|| ::||  ||   : ||
Db        411 MLATLVTYELMLL--QIDAKSHKG 432


RESULT 15
AAE30491
ID   AAE30491 standard; protein; 436 AA.
XX
AC   AAE30491;
XX
DT   24-FEB-2003  (first entry)
XX
DE   Fruit fly gustatory receptor protein, Gr61D1.
XX
KW   Fruit fly; gustatory receptor; Gr; odorant receptor; Or; pesticide;
KW   Gr61D1.
XX
OS   Drosophila melanogaster.
XX
PN   WO200268593-A2.
XX
PD   06-SEP-2002.
XX
PF   22-FEB-2002; 2002WO-US005414.
XX
PR   23-FEB-2001; 2001US-0271319P.
XX
PA   (UYCO ) UNIV COLUMBIA NEW YORK.
XX
```

PI Axel R, Scott K;
XX
DR WPI; 2002-698668/75.
XX
PT Novel nucleic acid encoding insect gustatory or odorant receptor protein
PT useful for identifying a compound which specifically binds to the
PT receptor for controlling a pest population in an area.
XX
PS Claim 3; Page 158-160; 264pp; English.
XX
CC The invention relates to an isolated nucleic acid encoding an insect
CC gustatory receptor (Gr) or odorant receptor (Or) protein which comprises
CC seven transmembrane domains and a C-terminal domain comprising
CC consecutive amino acids. The invention is useful for identifying a
CC compound which activates the insect receptor or inhibits the activity of
CC the insect receptor. The purified insect receptor protein is embedded in
CC a lipid bilayer. The invention is sprayed for combating ingestion of
CC crops by pest insects, combating disease-carrying insects in an area and
CC controlling a pest population in an area. The invention is useful for
CC detecting the presence of insect gustatory or odorant receptor and for
CC inhibiting the function of the receptor in humans or animals or in
CC biological fluids isolated from them. The invention is also useful for
CC identifying or isolating other insect receptors and for combating pest
CC nuisances. The present sequence is fruit fly gustatory receptor (Gr)
CC protein
XX
SQ Sequence 436 AA;

Query Match 4.9%; Score 129; DB 5; Length 436;
Best Local Similarity 21.2%; Pred. No. 2.9e-05;
Matches 94; Conservative 70; Mismatches 172; Indels 108; Gaps 22;

```
Qy         85 ASNNPEFKP-SVFYRNIDPINWFLRIIGVLP---IVRHGPARAKFEMNSASFIYSVVFFV 140
              |    ||:    ||  | |     :  |::|   |    |   ||:: |     : :| :
Db         30 AQGGLEFEQLDTFYGAIRPYLCVAQFFGIMPLSNIRSRDPQDVKFKVRSIGLAVTGLFLL 89

Qy        141 L--LACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYE----ARKIA 194
              |  :   ||  ||     :    |   : ::  :||:      :  : |      ||  :
Db         90 LGGMKTLVG--AN-----ILFTEGLNAKNIVGLVFLI------VGMVNWLNFVGFARSWS 136

Qy        195 KLFNDWDDFEVLY----YQISGHSLPLKLRQKAVYIAIVLPILSV--------------- 235
               :   |   ::|     |:    |    || |   :|: : :|:|
Db        137 HIMLPWSSVDILMLFPPYKRGKRS----LRSKVNVLALSVVVLAVGDHMLYYASGYCSYS 192

Qy        236 LSVVITHVTMSDL-----------NINQVVPYCILDNLTAM-LGAW--------WFLICE 275
              : ::  |   | :           :|  ::|:    |: :|  | |         |  :
Db        193 MHILQCHTNHSRITFGLYLEKEFSDIMFIMPF----NIFSMCYGFWLNGAFTFLWNFMDI 248

Qy        276 AMSITAHLLAERFQKALKHIG-------PAAMVADYRVLWLRLSKLTRDTGNALCYTFVF 328
               : :|:  ||:|||:    :|       | |:  | |   :|| :|
Db        249 FIVMTSIGLAQRFQQFAARVGALEGRHVPEALWYDIRRDHIRLCELA-------SLVEAS 301

Qy        329 MSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALW--------NIGLLFYICDEAHYA 380
              ||  :|     ::| : :|    |      :     |           |:|    : | |
Db        302 MSNIVFVSCANNVYVICNQALAIFTKLRHPINYVYFWYSLIFLLARTSLVFMTASKIHDA 361

Qy        381 SVNVRTNFQKKLLMVELNWMNSDAQT-EINMFLRATEMNPSTINCGGF--FDVNRTLFKG 437
              |:         | :  |  : ||  | |:  |  | ::    :   |:  | : |    |
Db        362 SL---------LPLRSLYLVPSDGWTQEVQRF--ADQLTSEFVGLSGYRLFCLTRKSLFG 410

Qy        438 LLTTMVTYLVVLLQFQISIPTDKG 461
              :| |:||| ::||  ||   : ||
Db        411 MLATLVTYELMLL--QIDAKSHKG 432
```

Search completed: May 23, 2005, 19:22:13
Job time : 175 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 23, 2005, 19:15:31 ; Search time 43 Seconds
(without alignments)
888.845 Million cell updates/sec

Title: US-10-081-816-12
Perfect score: 2645
Sequence: 1 MRPSGEKVVKGHGQGNSGHS.....TTTVGTTLPPPIMKLKGRKG 512

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	109	4.1	106	4	US-09-270-767-42643	Sequence 42643, A
2	107	4.0	404	4	US-09-270-767-33959	Sequence 33959, A
3	107	4.0	404	4	US-09-270-767-49176	Sequence 49176, A
4	104.5	4.0	323	3	US-09-134-001C-5200	Sequence 5200, Ap
5	104	3.9	363	4	US-09-491-577-74	Sequence 74, Appl
6	102	3.9	907	4	US-09-198-452A-306	Sequence 306, App
7	102	3.9	928	4	US-09-438-185A-295	Sequence 295, App
8	101	3.8	239	4	US-09-710-279-1696	Sequence 1696, Ap
9	101	3.8	395	4	US-09-491-577-98	Sequence 98, Appl
10	100.5	3.8	518	3	US-09-134-001C-4744	Sequence 4744, Ap
11	99.5	3.8	1989	4	US-09-949-016-10076	Sequence 10076, A
12	99	3.7	1900	4	US-09-538-092-564	Sequence 564, App
13	98.5	3.7	656	4	US-09-489-039A-8212	Sequence 8212, Ap
14	98	3.7	360	1	US-08-671-525B-6	Sequence 6, Appli
15	98	3.7	360	1	US-08-672-109B-6	Sequence 6, Appli
16	98	3.7	360	2	US-08-842-045-6	Sequence 6, Appli
17	98	3.7	360	2	US-08-842-238-6	Sequence 6, Appli
18	98	3.7	360	2	US-08-780-749A-1	Sequence 1, Appli
19	98	3.7	360	3	US-08-629-335B-6	Sequence 6, Appli
20	98	3.7	360	3	US-08-870-511-1	Sequence 1, Appli
21	98	3.7	360	4	US-09-709-066-4	Sequence 4, Appli
22	98	3.7	600	4	US-09-134-000C-5694	Sequence 5694, Ap
23	97.5	3.7	459	3	US-09-097-889-22	Sequence 22, Appl
24	97.5	3.7	459	4	US-09-098-079-22	Sequence 22, Appl
25	97.5	3.7	486	3	US-09-134-001C-3593	Sequence 3593, Ap
26	97	3.7	309	4	US-09-328-352-7637	Sequence 7637, Ap
27	97	3.7	391	1	US-07-816-283-2	Sequence 2, Appli
28	97	3.7	391	1	US-08-417-103-2	Sequence 2, Appli
29	97	3.7	391	1	US-08-417-103-14	Sequence 14, Appl
30	97	3.7	391	4	US-09-826-509-569	Sequence 569, App
31	96.5	3.6	450	4	US-09-721-870-16	Sequence 16, Appl
32	96.5	3.6	551	4	US-10-327-189-7	Sequence 7, Appli
33	96.5	3.6	551	4	US-09-521-195B-1	Sequence 1, Appli
34	96.5	3.6	822	4	US-09-824-734-3	Sequence 3, Appli
35	96	3.6	808	3	US-09-134-001C-3105	Sequence 3105, Ap
36	95.5	3.6	209	4	US-09-248-796A-21116	Sequence 21116, A
37	95	3.6	391	3	US-08-120-601B-8	Sequence 8, Appli
38	95	3.6	436	4	US-09-543-681A-4395	Sequence 4395, Ap
39	94.5	3.6	1116	4	US-09-543-681A-4379	Sequence 4379, Ap
40	94	3.6	411	4	US-09-489-039A-7912	Sequence 7912, Ap
41	94	3.6	520	4	US-09-252-991A-17058	Sequence 17058, A
42	93.5	3.5	551	4	US-10-327-189-38	Sequence 38, Appl
43	93	3.5	391	1	US-07-816-283-4	Sequence 4, Appli
44	93	3.5	391	1	US-08-417-103-4	Sequence 4, Appli
45	93	3.5	1058	4	US-09-328-352-4276	Sequence 4276, Ap

ALIGNMENTS

RESULT 1
US-09-270-767-42643
; Sequence 42643, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42643
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-42643

Query Match 4.1%; Score 109; DB 4; Length 106;
Best Local Similarity 45.1%; Pred. No. 0.0012;
Matches 23; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

Qy 405 QTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQIS 455
Db 46 QKEVEMLLVAITKNPPIMNLDGYANINRELITTNISFMATYLVVLLQFKIT 96

RESULT 2
US-09-270-767-33959
; Sequence 33959, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33959
; LENGTH: 404
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-33959

Query Match 4.0%; Score 107; DB 4; Length 404;
Best Local Similarity 18.2%; Pred. No. 0.016;
Matches 75; Conservative 85; Mismatches 155; Indels 96; Gaps 17;

Qy 106 FLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFE 165

```
Db         23 FMTVFGLLANRYRAGRRERFRFSKANLAFASLWAIAFSLVYG----------RQIYKEYQ 72

Qy        166 E-------AVIAYLFL------VNILPIMIIPILWYEARKIAKLFNDWDDFEVLYYQISG 212
Db         73 EGQINLKDATTLYSYMNITVAVINYVSQMII------SDHVAKVLSKVPFFDTL------ 120

Qy        213 HSLPLKLRQKAVYIAIVLPILSVLSVVIT------------HVTMS---------DLNIN 251
Db        121 --KEFRLDSRSLYISIVLALVKTVAFPLTIEVAFILQQRRQHPEMSLIWTLYRLFPLIIS 178

Qy        252 QVVPYCILDNLTAMLGAWWFL--ICEAMSITAHLLAERFQKAL----KHIGPAAMVAD-- 303
Db        179 NFLNNCYFGAMVVVKEILYALNRRLEAQLQEVNLLQRKDQLKLYTKYYRMQRFCALADEL 238

Qy        304 ------YRVLWLRLSKLTRDTG-----NALCY----TFVFMSLYLFFIITLSIYGLMSQL 348
Db        239 DQLAYRYRLIYVHSGKYLTPMSLSMILSLICHLLGITVGFYSLYYAIADTL----IMGKP 294

Qy        349 SEGFG--IKDIGLTITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDA-- 404
Db        295 YDGLGSLINLVFLSI-SLAEITLLTHLCNHLLVAT------RRSAVILQEMNLQHADSRY 347

Qy        405 QTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQIS 455
Db        348 RQAVHGFTLLVTVTKYQIKPLGLYELDMRLISNVFSAVASFLLILVQADLS 398


RESULT 3
US-09-270-767-49176
; Sequence 49176, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49176
;   LENGTH: 404
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-270-767-49176

  Query Match             4.0%;  Score 107;  DB 4;  Length 404;
  Best Local Similarity   18.2%;  Pred. No. 0.016;
  Matches   75;  Conservative   85;  Mismatches  155;  Indels   96;  Gaps   17;

Qy        106 FLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFE 165
Db         23 FMTVFGLLANRYRAGRRERFRFSKANLAFASLWAIAFSLVYG----------RQIYKEYQ 72

Qy        166 E-------AVIAYLFL------VNILPIMIIPILWYEARKIAKLFNDWDDFEVLYYQISG 212
Db         73 EGQINLKDATTLYSYMNITVAVINYVSQMII------SDHVAKVLSKVPFFDTL------ 120

Qy        213 HSLPLKLRQKAVYIAIVLPILSVLSVVIT------------HVTMS---------DLNIN 251
Db        121 --KEFRLDSRSLYISIVLALVKTVAFPLTIEVAFILQQRRQHPEMSLIWTLYRLFPLIIS 178

Qy        252 QVVPYCILDNLTAMLGAWWFL--ICEAMSITAHLLAERFQKAL----KHIGPAAMVAD-- 303
Db        179 NFLNNCYFGAMVVVKEILYALNRRLEAQLQEVNLLQRKDQLKLYTKYYRMQRFCALADEL 238

Qy        304 ------YRVLWLRLSKLTRDTG-----NALCY----TFVFMSLYLFFIITLSIYGLMSQL 348
Db        239 DQLAYRYRLIYVHSGKYLTPMSLSMILSLICHLLGITVGFYSLYYAIADTL----IMGKP 294

Qy        349 SEGFG--IKDIGLTITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDA-- 404
Db        295 YDGLGSLINLVFLSI-SLAEITLLTHLCNHLLVAT------RRSAVILQEMNLQHADSRY 347

Qy        405 QTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQIS 455
Db        348 RQAVHGFTLLVTVTKYQIKPLGLYELDMRLISNVFSAVASFLLILVQADLS 398


RESULT 4
US-09-134-001C-5200
; Sequence 5200, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5200
;   LENGTH: 323
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5200

  Query Match             4.0%;  Score 104.5;  DB 3;  Length 323;
  Best Local Similarity   22.7%;  Pred. No. 0.02;
  Matches   58;  Conservative   34;  Mismatches   79;  Indels   85;  Gaps   10;

Qy        271 FLICEAMSITAHLLAERFQKALKHIGPAAMVADY-RVLWLRLSKLTRDTGNALCYTFVFM 329
Db         66 FIINDEAAYQVHLSNSFEPMSMNHL---LGTDDYGRDLFSRLVVGSRAT------LFVTL 116

Qy        330 SLYLFFIITLSIYGLMSQLSEGF------GIKDIGLTI--------------TALWNIGL 369
Db        117 LTLLFTVVVGVPLGLLAGYKKGWIDTIIMRIIDIGLSIPEFVIMIALASFFHPSLWNLVI 176

Qy        370 LFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFD 429
Db        177 AITI----------------------IKWMN---YTRVTRGIVNTEMNQSYIQMAQFFN 210

Qy        430 VN--RTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLDNDMSLMG 487
Db        211 VSTLNILFKHLLPKVLPSIFVIM--------------------IVDF-------GKIILY 243

Qy        488 ASTLSTTTVGTTLPPP 503
Db        244 ISSLSFLGLGAQPPSP 259


RESULT 5
US-09-491-577-74
; Sequence 74, Application US/09491577
; Patent No. 6610511
; GENERAL INFORMATION:
;  APPLICANT: Yale University
;  APPLICANT: Carlson, John R.
;  APPLICANT: Kim, Hunhyong
;  APPLICANT: Clyne, Peter J.
;  APPLICANT: Warr, Coral G.
;  TITLE OF INVENTION: No. 6610511el Family of Odorant Receptor Genes in Drosophila
;  FILE REFERENCE: 44574-5061-US
;  CURRENT APPLICATION NUMBER: US/09/491,577
;  CURRENT FILING DATE: 2000-01-25
;  EARLIER APPLICATION NUMBER: US 60/117,132
;  EARLIER FILING DATE: 1999-01-25
;  NUMBER OF SEQ ID NOS: 112
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 74
;   LENGTH: 363
```

```
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-491-577-74

  Query Match             3.9%;  Score 104;  DB 4;  Length 363;
  Best Local Similarity   22.9%;  Pred. No. 0.028;
  Matches   77;  Conservative   43;  Mismatches  124;  Indels   92;  Gaps   17;

Qy        122 RAKFEMNSASF---IYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLF-LVNI 177
              |:| |::   |    :: ||      |  | :              ||   ||  || : |
Db        109 RSKQEVDMWRPEHRRFNRVFMFYCLCSAGVI------------PF--IVIQPLFDIPNR 153

Qy        178 LPIMIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLS 237
              ||      :|           ||    :|:|             :| |  :||
Db        154 LPFW----MWTPF--------DWQQPVLLWYAF------------IYQATTIPI----- 183

Qy        238 VVITHVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGP 297
                :|||  :|                    |:|:   :|:   :| :|  | |:|
Db        184 ACACNVTMDAVN------------------WYLMLH-LSLCLRMLGQRLSK-LQHDD- 220

Qy        298 AAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDI 357
                | |  :| |  | :          :|:|   |  |       |:| |   | |  :
Db        221 ----KDLREKFLELIHLHQRLKQQALSIEIFISKSTFTQI------LVSSLIICFTIYSM 270

Qy        358 GLTITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNW--MNSDAQTEINMFLRAT 415
              : : |:   :|  |   |  | |: |:       |   :|  ||   :  : ||:
Db        271 QMYLVAMIMQVMLPTIYGNAVIDSANMLTD-----SMYNSDWPDMNCRMRRLVLMFM--V 323

Qy        416 EMN-PSTINCGGFFDVNRTLFKGLLTTMVTYLVVLL 450
              :| | |:  |||| :   ||   :    : | :||
Db        324 YLNRPVTLKAGGFFHIGLPLFTKTMNQAYSLLALLL 359


RESULT 6
US-09-198-452A-306
; Sequence 306, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
;  APPLICANT: Griffais, R.
;  TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragmente
;  TITLE OF INVENTION:  thereof and uses thereof, in particular for the diagnosis, preve
;  TITLE OF INVENTION:  and treatment of infection
;  FILE REFERENCE: 9710-003-999
;  CURRENT APPLICATION NUMBER: US/09/198,452A
;  CURRENT FILING DATE:  1998-11-24
;  NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 306
;   LENGTH: 907
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
US-09-198-452A-306

  Query Match             3.9%;  Score 102;  DB 4;  Length 907;
  Best Local Similarity   18.8%;  Pred. No. 0.18;
  Matches   90;  Conservative   71;  Mismatches  149;  Indels  168;  Gaps   25;

Qy         88 NPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGPARAKFEMNSAS---FIYSVVFFVLLAC 144
              :|    |  |||    : | | |:        |      | :        |:: : |:  |
Db        112 DPPRSPLFFYR---IVIWSLTILSYTSF--WGFVDQFFNLQDGKRHFCIFNAIIFLGDAI 166

Qy        145 YVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFE 204
                | :| : :| :          :   | |     ::  ||:| :: :  | :| |
Db        167 GSGIIA-SLVHTI---------GIQGILILFTAALVLTFPIVFYVSKSLKSLSDDHD--- 213

Qy        205 VLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTMSDLNINQVVPYCILDNLTA 264
                :  :||  ||                                  :: :  |  |  |
Db        214 --LFIDTGHPPPL--------------------------------SKALKLCFYDKYTF 238

Qy        265 MLGAWWFLICEAMSITA---HLLAERFQKALK-------HIGPAAMVADYRVLWLRL--- 311
              |  ::||: : ::|     :|      | | |       |||  :       ||: |
Db        239 YLLCFYFLM-QLLAIATEFNYLKIFEIQFASKEEFELVAHIGKCS-------LWISLGNM 290

Qy        312 -------SKLTRDTG--NALCYT-FVFMSLYLF--FIITLSIYGLMSQLSEG-------- 351
                    |:: :  |  | : :    |:||:||  |  ||||  |   : ||
Db        291 CFALFAYSRIVKRLGVNNIILFAPLCFLSLFLFWTFKTTLSIAVLAMVVREGVTYALDDN 350

Qy        352 ------FG---------------IKDIGLTITALWN--------------IGLL--FY 372
                   :|                |: ||: :   |:              | |:
Db        351 NLQLLIYGVPNKIRNQIRIVVESFIEPIGMLV---WSLVCFLSSQQYVFCLIISLIATIL 407

Qy        373 ICDEAHYASVNVRTNFQKKLLMVE---LNWMNS---DAQTEINMFLRATEMNPSTINCGG 426
              :|    | :  :  |   : | :     :|: |     : :: :|| |   :||  :
Db        408 VCLVRSYYAKAILKNLSAQALQLTRSMQDWIKSMTVKQKRQVELFLLAHLKHPSERHQTF 467

Qy        427 FFD-----VNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSL 479
              |       :|::   ||  |          ::|:|          | :  :: |  ||
Db        468 AFQHLLNLASRSVLPSLLAHM---------NKLSLP---------NKLKTIEMVKSSL 507


RESULT 7
US-09-438-185A-295
; Sequence 295, Application US/09438185A
; Patent No. 6822071
; GENERAL INFORMATION:
;  APPLICANT: Stephens, Richard
;  APPLICANT:  Mitchell, Wayne
;  APPLICANT:  Kalman, Sue
;  APPLICANT:  Davis, Ronald
;  APPLICANT:  The Regents of the University of California
;  TITLE OF INVENTION: Chlamydia Pneumoniae Genome Sequence
;  FILE REFERENCE: 018941-000411US
;  CURRENT APPLICATION NUMBER: US/09/438,185A
;  CURRENT FILING DATE:  2002-03-13
;  PRIOR APPLICATION NUMBER: US 60/108,279
;  PRIOR FILING DATE: 1998-11-12
;  PRIOR APPLICATION NUMBER: US 60/128,606
;  PRIOR FILING DATE: 1999-04-08
;  NUMBER OF SEQ ID NOS: 1074
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 295
;   LENGTH: 928
;   TYPE: PRT
;   ORGANISM: Chlamydia pneumoniae
;   FEATURE:
;   OTHER INFORMATION: CPn0293
US-09-438-185A-295

  Query Match             3.9%;  Score 102;  DB 4;  Length 928;
  Best Local Similarity   18.8%;  Pred. No. 0.19;
  Matches   90;  Conservative   71;  Mismatches  149;  Indels  168;  Gaps   25;

Qy         88 NPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGPARAKFEMNSAS---FIYSVVFFVLLAC 144
              :|    |  |||    : | | |:        |      | :        |:: : |:  |
Db        112 DPPRSPLFFYR---IVIWSLTILSYTSF--WGFVDQFFNLQDGKRHFCIFNAIIFLGDAI 166

Qy        145 YVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFE 204
                | :| : :| :          :   | |     ::  ||:| :: :  | :| |
Db        167 GSGIIA-SLVHTI---------GIQGILILFTAALVLTFPIVFYVSKSLKSLSDDHD--- 213

Qy        205 VLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTMSDLNINQVVPYCILDNLTA 264
                :  :||  ||                                  :: :  |  |  |
Db        214 --LFIDTGHPPPL--------------------------------SKALKLCFYDKYTF 238

Qy        265 MLGAWWFLICEAMSITA---HLLAERFQKALK-------HIGPAAMVADYRVLWLRL--- 311
              |  ::||: : ::|     :|      | | |       |||  :       ||: |
Db        239 YLLCFYFLM-QLLAIATEFNYLKIFEIQFASKEEFELVAHIGKCS-------LWISLGNM 290

Qy        312 -------SKLTRDTG--NALCYT-FVFMSLYLF--FIITLSIYGLMSQLSEG-------- 351
                    |:: :  |  | : :    |:||:||  |  ||||  |   : ||
Db        291 CFALFAYSRIVKRLGVNNIILFAPLCFLSLFLFWTFKTTLSIAVLAMVVREGVTYALDDN 350

Qy        352 ------FG---------------IKDIGLTITALWN--------------IGLL--FY 372
                   :|                |: ||: :   |:              | |:
Db        351 NLQLLIYGVPNKIRNQIRIVVESFIEPIGMLV---WSLICFLSSQQYVFCLIISLIATIL 407

Qy        373 ICDEAHYASVNVRTNFQKKLLMVE---LNWMNS---DAQTEINMFLRATEMNPSTINCGG 426
```

```
Db        408 VCLVRSYYAKAILKNLSAQALQLTRSMQDWIKSMTVKQKRQVELFLLAHLKHPSERHQTF 467

Qy        427 FFD-----VNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSL 479
Db        468 AFQHLLNLASRSVLPSLLAHM---------NKLSLP---------NKLKTIEMVKSSL 507


RESULT 8
US-09-710-279-1696
; Sequence 1696, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;  APPLICANT: KIMMERLY, WILLIAM JOHN
;  TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;  FILE REFERENCE: PU3480US
;  CURRENT APPLICATION NUMBER: US/09/710,279
;  CURRENT FILING DATE:  2000-11-09
;  PRIOR APPLICATION NUMBER: 60/164,258
;  PRIOR FILING DATE: 1999-11-09
;  NUMBER OF SEQ ID NOS: 4472
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1696
;   LENGTH: 239
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: amino acid sequence
US-09-710-279-1696

  Query Match             3.8%;  Score 101;  DB 4;  Length 239;
  Best Local Similarity   23.7%;  Pred. No. 0.03;
  Matches   53;  Conservative   28;  Mismatches   61;  Indels   82;  Gaps    9;

Qy        303 DY-RVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGF------GIK 355
Db         11 DYGRDLFSRLVVGSRAT------LFVTLLTLLFTVVVGVPLGLLAGYKKGWIDTIIMRII 64

Qy        356 DIGLTI--------------TALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMN 401
Db         65 DIGLSIPEFVIMIALASFFHPSLWNLVIAITI----------------------IKWMN 101

Qy        402 SDAQTEINMFLRATEMNPSTINCGGFFDVN--RTLFKGLLTTMVTYLVVLLQFQISIPTD 459
Db        102 ---YTRVTRGIVNTEMNQSYIQMAQFFNVSTLNILFKHLLPKVLPSIFVIM--------- 149

Qy        460 KGDSEGANNITVVDFVMDSLDNDMSLMGASTLSTTTVGTTLPPP 503
Db        150 -----------IVDF-------GKIILYISSLSFLGLGAQPPSP 175


RESULT 9
US-09-491-577-98
; Sequence 98, Application US/09491577
; Patent No. 6610511
; GENERAL INFORMATION:
;  APPLICANT: Yale University
;  APPLICANT: Carlson, John R.
;  APPLICANT: Kim, Hunhyong
;  APPLICANT: Clyne, Peter J.
;  APPLICANT: Warr, Coral G.
;  TITLE OF INVENTION: No. 6610511el Family of Odorant Receptor Genes in Drosophila
;  FILE REFERENCE: 44574-5061-US
;  CURRENT APPLICATION NUMBER: US/09/491,577
;  CURRENT FILING DATE: 2000-01-25
;  EARLIER APPLICATION NUMBER: US 60/117,132
;  EARLIER FILING DATE: 1999-01-25
;  NUMBER OF SEQ ID NOS: 112
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 98
;   LENGTH: 395
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-09-491-577-98

  Query Match             3.8%;  Score 101;  DB 4;  Length 395;
  Best Local Similarity   18.9%;  Pred. No. 0.066;
  Matches   56;  Conservative   58;  Mismatches  112;  Indels   70;  Gaps   12;

Qy        222 KAVYIAIVL---PILSV-------LSVVITH-------------VTMSDLNINQVVPY-- 256
Db         97 KSFYIELLQSDDPIINILVKETTRLSVLISRINLLMGCCTCIGFVTYPIFGSERVLPYGM 156

Qy        257 -----------------CILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKH-IGP 297
Db        157 YLPTIDEYKYASPYYEIFFVIQAIMAPMGCCMYIPYTNMVVTFTLFAILMCRVLQHKLRS 216

Qy        298 AAMVADYRV----LW-----LRLSKLTRDTGNALCYTFVFMSLYLFF--IITLSIYGLMS 346
Db        217 LEKLKNEQVRGEIIWCIKYQLKLSGFV-DSMNAL-NTHLHLVEFLCFGAMLCVLLFSLII 274

Qy        347 QLSEGFGIKDIGLTITALWNIGLLFYICDEAHYA------------SVNVRTNFQKKLLM 394
Db        275 AQTIAQTVIVIAYMVMIFANSVVLYYVANELYFQVRVVQFSFKFLYKYGILQSFDIAIAA 334

Qy        395 VELNWMNSDAQTEINM-FLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVL 449
Db        335 YESNWMDFDVDTQKTLKFLIMRSQKPLATLVGGTYPMNLKMLQSLLNAIYSFFTLL 390


RESULT 10
US-09-134-001C-4744
; Sequence 4744, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4744
;   LENGTH: 518
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4744

  Query Match             3.8%;  Score 100.5;  DB 3;  Length 518;
  Best Local Similarity   19.9%;  Pred. No. 0.11;
  Matches   94;  Conservative   63;  Mismatches  139;  Indels  177;  Gaps   21;

Qy         93 PSVFYRNIDPINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANN 152
Db        130 PDFFKNRLDDKKNIIKIISGLII-------------------VVFFTLYT-HSGFVSGG 168

Qy        153 RIHIVRSLSGPFEEA---------------VIAYLFLVNILP----------IMIIPILW 187
Db        169 KL---------FESAFGLNYHAGLLIVAIIVIFYTFFGGYLAVSITDFFQGVIMLIAMVM 219

Qy        188 YEARKIAKLFNDWDDFE-----------------------------------VLYYQ 209
Db        220 VPIVALLKL-NGWDTFHDIAQMKPTNLDLFRGTTVLGIVSLFSWGLGYFGQPHIIVRFMS 278

Qy        210 ISGHSLPLKLRQKAV-YIAIVLPILSVLSVVIT--------HVTMSD-----LNINQVVP 255
Db        279 IKSHKLLPKARRLGISWMAVGL--LGAIGVGLTGISFISERHIKLEDPETLFIVMSQILF 336

Qy        256 YCILDN--LTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSK 313
Db        337 HPLVGGFLLAAILAAIMSTISSQLLVTSSSLTEDFYKLIRGSDKASSHQKEFVLIGRLSV 396
```

```
Qy        314 LTRDTGNALCYTFVFMSLYLFFIITL------SIYGLMSQLSEGFGIKDIGLTITALWNI 367
              |                 |     ||:      :|  |:      |||     | : :|
Db        397 L---------------LVAIVAITIAWHPNDTILNLVGNAWAGFGAAFSPLVLYSL--- 437

Qy        368 GLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGF 427
              | |      :::         :::| ::|:   |   || |     |: |       ||
Db        438 ----YWKDLTRAGAISGMV-AGAVVVIVWISWIKPLA--TINAFFGMYEIIP------GF 484

Qy        428 FDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLD 480
                     ::: ::||:|  |        | | |           |:|:::|:
Db        485 ----------IVSVLITYIVSKL-------TKKPD----------DYVIENLN 510


RESULT 11
US-09-949-016-10076
; Sequence 10076, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE:  2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10076
;   LENGTH: 1989
;   TYPE: PRT
;   ORGANISM: Human
US-09-949-016-10076

  Query Match             3.8%;  Score 99.5;  DB 4;  Length 1989;
  Best Local Similarity   19.8%;  Pred. No. 1.1;
  Matches  102;  Conservative   80;  Mismatches  177;  Indels  155;  Gaps   26;

Qy         57 KYSIGLAERLDADYEAPPLDRKKSSDS-TASNNPEFKPSVFYRNIDPIN-WFLRII---- 110
              ||      |  :|   :| |   :     || | :      | | :   :| ::
Db       1361 KYHYCFNETSEIRFEIEDVNNKTECEKLMEGNNTEIRWKNVKINFDNVGAGYLALLQVAT 1420

Qy        111 --GVLPIV------RHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSG 162
                | : |:      |    : |:| |   :|| |:| :                    |
Db       1421 FKGWMDIMYAAVDSRKPDEQPKYEDNIYMYIYFVIFIIF--------------------G 1460

Qy        163 PFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQK 222
              |        | :|:   :||                 |:|     :  | : :   ||
Db       1461 SF--------FTLNLFIGVII-----------------DNFNQQKKKFGGQDIFMTEEQK 1495

Qy        223 AVYIA------------IVLPILSVLSVVITHVTMSDLNINQVVPYCILDNLTAMLG--- 267
              | |              |  |:  :  :|   ||    :|  ::  |:  |:  |:
Db       1496 KYYNAMKKLGSKKPQKPIPRPLNKIQGIVFDFVTQQAFDIVIMMLICL--NMVTMMVETD 1553

Qy        268 ---------AWWFLICEAMSITAHLLAERFQKALKH----IGPAAMVADYRVLWLRLSKL 314
                      :|  :   :   |   : : |  ||:|     ||     : |: |: | :
Db       1554 TQSKQMENILYWINLVFVIFFTCECVLKMF--ALRHYYFTIG--WNIFDFVVVILSI--- 1606

Qy        315 TRDTGNALCYTF--VFMSLYLFFIITLSIYGLMSQLSEGF-GIKD----IGLTITALWNI 367
                |  |        |:|  || :| |:  | : :| :|  ||:      : ::: ||:||
Db       1607 ---VGMFLADIIEKYFVSPTLFRVIRLARIGRILRLIKGAKGIRTLLFALMMSLPALFNI 1663

Qy        368 GLLFYI------------CDEAHYASVNVRTNFQ-----------------KKLLMVE 396
              ||| ::                | | ::   ||:                    ||:
Db       1664 GLLLFLVMFIFSIFGMSNFAYVKHEAGIDDMFNFETFGNSMICLFQITTSAGWDGLLLPI 1723

Qy        397 LNW---MNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQ 453
              ||      : | :   : | :    |||    | || |:  :   |:   : |: ::|:
Db       1724 LNRPPDCSLDKEHPGSGF-KGDCGNPSV---GIFFFVSYIIISFLIVVNM-YIAIILE-N 1777

Qy        454 ISIPTDKGDSEGANNITVVDF-----VMDSLDND 482
                |: |:    | |: ::  ||      : :  | |
Db       1778 FSVATE----ESADPLSEDDFETFYEIWEKFDPD 1807


RESULT 12
US-09-538-092-564
; Sequence 564, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
;  APPLICANT: Giot, Loic
;  APPLICANT:  Mansfield, Traci A.
;  TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
;  FILE REFERENCE: 15966-542
;  CURRENT APPLICATION NUMBER: US/09/538,092
;  CURRENT FILING DATE:  2000-03-29
;  PRIOR APPLICATION NUMBER: 60/127,352
;  PRIOR FILING DATE: 1999-04-01
;  PRIOR APPLICATION NUMBER: 60/178,965
;  PRIOR FILING DATE: 2000-02-01
;  NUMBER OF SEQ ID NOS: 1387
;  SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 564
;   LENGTH: 1900
;   TYPE: PRT
;   ORGANISM: Saccharomyces cerevisiae
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (0)...(0)
;   OTHER INFORMATION: Polypeptide Accession Number YLR305C
US-09-538-092-564

  Query Match             3.7%;  Score 99;  DB 4;  Length 1900;
  Best Local Similarity   19.2%;  Pred. No. 1.2;
  Matches   90;  Conservative   66;  Mismatches  154;  Indels  158;  Gaps   20;

Qy         70 YEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGPAR------- 122
              | :|||     :    :|| |                    : :  |:| | :
Db        727 YNSPPL-----ASEFPANNKELS----------------LEMNTILRRGSSNENIKQQK 764

Qy        123 -----------AKFEMNSAS---FIYSVVFFVLLACYVGYVANNRIHI--VRSLSGPFEE 166
                        ::   |:|   |: : |    :|  |  :   ::      |||  |:
Db        765 QQITEYFNTNIVQYRTTSSSKIMFLAAAVLLETIRCEAGDCSKTLLYFSDPSILSGSIEK 824

Qy        167 AVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDD----FEVLYYQISGHSLPLKLR-- 220
              :       :| : :|       || |:|    :|     ::: |::   | |  |
Db        825 CIA-------VLSVSMI-------RKYARLIQKGNDAIFNSKMIAQQLNNLLLCLSHREP 870

Qy        221 --QKAVYIAIVLPILSVLSVVITHVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMS 278
                | : |  : | |: | :  |:::           | :|| |||:  :        :
Db        871 TLQDAAFHACEIFIRSIPSSLCHHLSL----------YTLLDMLTALFDS-------ILD 913

Qy        279 ITAHLLAERFQKALKHIGPAAMV---ADYRVL----------WLR--LSKLTRDTGNAL 322
                ||     |::  |||      :|   : :|              |:|  |::  :||
Db        914 SEAHKFEPRYEFKLKHSKTTILVPSSSSWRATTLSRLHKSAKEWVRILLNRSNQDT---- 969

Qy        323 CYTFVFMSLYLFFIITLSIYGLMSQLSEG--FGIKDIGLTITALWNIGLLFYICDEAHYA 380
                  |   :|  |  |  :: :  |  | :   || : |   :  | |   |
Db        970 ------KILLQSYISDLGEYSRLNSVEFGVSFAMDMAGLILPADKELSRLTYYGPE---- 1019

Qy        381 SVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLT 440
                                                 |:||:  ||  ::    | |
Db       1020 ------------------------------------KPNTIS--GFISLHSWRSKYLFD 1040

Qy        441 TMVTYLVVLLQFQISIPTD--KGDSEGANNITVVDFVMDSLDNDMSLM 486
              | :|     :: || | |   : :    | |   | | | ||   :|:
Db       1041 TAITSSPEDIKRQIGISTQNIRKNLTLGNKIITKD-VTDFLDMATALL 1087


RESULT 13
US-09-489-039A-8212
```

```
; Sequence 8212, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT:  Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 8212
;   LENGTH: 656
;   TYPE: PRT
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8212

  Query Match             3.7%;  Score 98.5;  DB 4;  Length 656;
  Best Local Similarity   21.2%;  Pred. No. 0.26;
  Matches  118;  Conservative   67;  Mismatches  206;  Indels  165;  Gaps   28;

Qy         32 KGDAVFLNAKP-----LNSANAQAYLYGVRKYSIGLAERLDADY--EAPPLDRKKSSDST 84
Db         75 KADAEGLKANPGVIMVVESGGQFQVVIGNHVHDVWLAVRQEAGLSDDSEPVAEEKAA--- 131

Qy         85 ASNNPEFKPSVFYRNIDPINW-FLRIIGVLPIV-----------------RHGPARAKF 125
Db        132 -------KGSVLSQLIDIISGIFTPFIGVMAATGLLKGLLALAVTCGWLTPEQGTYKIWF 184

Qy        126 EMNSASFIYSVVFFVLLACYVGYVANNRI----HIVRSLSGPFEEAVIAYLFLVNILP-- 179
Db        185 AASDALFF----FFPL---FLGYTAGKKFGGNPFISMVIGGALTHPLMIQAFEASQAPGA 237

Qy        180 ----IMIIPILW--YEARKIAKLFNDWDDFEVLYYQISGHSLPLKLR---QKAVYIAIVL 230
Db        238 AVEHFLGIPVTFINYSSSVIPIILASW--VSCWLERKSNALLPSSMKNFFSPAICLAVVV 295

Qy        231 PI-LSVLSVVITHVTMSDLNINQVVPYCILDNLT-AMLGAWWF-------------LICE 275
Db        296 PLTFLVIGPVATWLSHLLLANGYQFI-YAFAPWLAGAVLGAMWQVCVIFGLHWGLVPLMIN 354

Qy        276 AMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFF 335
Db        355 NMTVLGH------DSMLPIILPAVIAQVGAVLGIFLA--TRDARQRVLAGSAF-SAGLFG 405

Qy        336 IITLSIYGLMSQLSEGF------GIKDIGLTITAL-----WNIGLLFYICDEAHYASVNV 384
Db        406 ITEPAIYGLTLPLRRPFIFGCVAGA--IGGAITAFSNSYAYSFGL--------------- 448

Qy        385 RTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVT 444
Db        449 ----------------------PNIFFPAQMIPP-----GG---IDASVWGGLIGTGVA 477

Qy        445 Y-LVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLDNDMS-----------------LM 486
Db        478 FVLACVLTFFAGLPR---GSAAPGAVTVAPASANDILAPMSGSVIALEQVPDSTFASGLL 534

Qy        487 GASTLSTTTVGTTLPP 502
Db        535 GKGVAIIPAVGQVIAP 550


RESULT 14
US-08-671-525B-6
; Sequence 6, Application US/08671525B
; Patent No. 5703220
;  GENERAL INFORMATION:
;    APPLICANT:  Yamada, Tadataka
;    APPLICANT:  Gantz, Ira
;    TITLE OF INVENTION:  Genes Encoding Melanocortin Receptors
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  US
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/671,525B
;      FILING DATE:  June 27, 1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-000853DVB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810)641-1600
;      TELEFAX:  (810)641-0270
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  360 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-671-525B-6

  Query Match             3.7%;  Score 98;  DB 1;  Length 360;
  Best Local Similarity   21.3%;  Pred. No. 0.12;
  Matches   85;  Conservative   64;  Mismatches  172;  Indels   78;  Gaps   17;

Qy         32 KGDAVF---------------LNAKPLNSANAQAYLYGVRKYSIGLAERLDADYEAPPLD 76
Db          9 EGDFVFPVSSSSFLRTLLEPQLGSALLTAMNASCCLPSVQPTLPNGSEHLQAPFFS---- 64

Qy         77 RKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSV 136
Db         65 --NQSSSAFCEQVFIKPEIFL-SLGIVSLLENILVILAVVRNGNLHSPMYF----FLCSL 117

Qy        137 VFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKL 196
Db        118 AVADMLVSVSNALETIMIAIVHSDYLTFEDQFIQH--MDNIFDSMICISL------VASI 169

Qy        197 FN----DWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTMSDLNINQ 252
Db        170 CNLLAIAVDRYVTIFYALRYHSI-MTVRKALTLIVAIWVCCGVCGVVF--IVYSE---SK 223

Qy        253 VVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLS 312
Db        224 MVIVCLITMFFAMM-----LLMGTLYVHMFLFARLHVKRIAALPPADGVAPQQHSCMKGA 278

Qy        313 KLTRDTGNALCYTFVF----MSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIG 368
Db        279 V----TITILLGVFIFCWAPFFLHLVLIITCP------------TNPYCICYTAHFNTY 321

Qy        369 LLFYIC----DEAHYA--SVNVRTNFQKKLLMVELNWMN 401
Db        322 LVLIMCNSVIDPLIYAFRSLELRNTFRE--ILCGCNGMN 358


RESULT 15
US-08-672-109B-6
; Sequence 6, Application US/08672109B
; Patent No. 5710265
;  GENERAL INFORMATION:
;    APPLICANT:  Yamada, Tadataka
;    APPLICANT:  Gantz, Ira
;    TITLE OF INVENTION:  Genes Encoding Melanocortin Receptors
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  Harness, Dickey & Pierce, P.L.C.
;      STREET:  P.O. Box 828
;      CITY:  Bloomfield Hills
;      STATE:  MI
;      COUNTRY:  US
;      ZIP:  48303
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/672,109B
;      FILING DATE:  June 27, 1996
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Smith, DeAnn F.
;      REGISTRATION NUMBER:  36683
;      REFERENCE/DOCKET NUMBER:  2115-000853DVC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (810)641-1600
;      TELEFAX:  (810)641-0270
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  360 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-672-109B-6

  Query Match             3.7%;  Score 98;  DB 1;  Length 360;
  Best Local Similarity   21.3%;  Pred. No. 0.12;
  Matches   85;  Conservative   64;  Mismatches  172;  Indels   78;  Gaps   17;

Qy         32 KGDAVF---------------LNAKPLNSANAQAYLYGVRKYSIGLAERLDADYEAPPLD 76
              :|| ||               | :  | : ||   |  |:      :| | | : :
Db          9 EGDFVFPVSSSSFLRTLLEPQLGSALLTAMNASCCLPSVQPTLPNGSEHLQAPFFS---- 64

Qy         77 RKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSV 136
                | |         || :|  ::  ::    |: :| :||:|   :         |: |:
Db         65 --NQSSSAFCEQVFIKPEIFL-SLGIVSLLENILVILAVVRNGNLHSPMYF----FLCSL 117

Qy        137 VFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKL 196
                :|      :   | || |    ||:  | :  : ||   ||   |       :| :
Db        118 AVADMLVSVSNALETIMIAIVHSDYLTFEDQFIQH--MDNIFDSMICISL------VASI 169

Qy        197 FN----DWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTMSDLNINQ 252
               |       | :  ::| :  ||: : :|:    |  :     |  ||   :  |:   ::
Db        170 CNLLAIAVDRYVTIFYALRYHSI-MTVRKALTLIVAIWVCCGVCGVVF--IVYSE---SK 223

Qy        253 VVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLS 312
               :|  |::    ||:     |:   : :   | |    | :  : ||  ||  :   :: :
Db        224 MVIVCLITMFFAMM-----LLMGTLYVHMFLFARLHVKRIAALPPADGVAPQQHSCMKGA 278

Qy        313 KLTRDTGNALCYTFVF----MSLYLPFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIG 368
                  |   |   |:|     |:|  |||            :  || :|
Db        279 V----TITILLGVFIFCWAPFFLHLVLIITCP-------------TNPYCICYTAHFNTY 321

Qy        369 LLFYIC----DEAHYA--SVNVRTNFQKKLLMVELNWMN 401
              |:  :|    |   ||  |: :|  |::  ::   | ||
Db        322 LVLIMCNSVIDPLIYAFRSLELRNTFRE--ILCGCNGMN 358


Search completed: May 23, 2005, 19:26:45
Job time : 46 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 23, 2005, 19:19:25 ; Search time 135 Seconds
(without alignments)
1268.651 Million cell updates/sec

Title: US-10-081-816-12
Perfect score: 2645
Sequence: 1 MRPSGEKVVKGHGQGNSGHS.....TTTVGTTLPPPIMKLKGRKG 512

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1434725 seqs, 334507595 residues

Total number of hits satisfying chosen parameters: 1434725

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2645	100.0	512	14	US-10-081-816-12	Sequence 12, Appl
2	390	14.7	364	14	US-10-081-816-24	Sequence 24, Appl
3	390	14.7	364	15	US-10-447-328-10	Sequence 10, Appl
4	169	6.4	33	14	US-10-081-816-67	Sequence 67, Appl
5	160	6.0	433	15	US-10-447-328-80	Sequence 80, Appl
6	135.5	5.1	369	14	US-10-081-816-29	Sequence 29, Appl
7	132.5	5.0	736	14	US-10-081-816-17	Sequence 17, Appl
8	130.5	4.9	372	14	US-10-081-816-30	Sequence 30, Appl
9	130.5	4.9	372	15	US-10-447-328-32	Sequence 32, Appl
10	129	4.9	436	14	US-10-081-816-11	Sequence 11, Appl
11	127	4.8	348	15	US-10-447-328-30	Sequence 30, Appl
12	122	4.6	410	14	US-10-081-816-1	Sequence 1, Appli
13	122	4.6	414	15	US-10-447-328-2	Sequence 2, Appli
14	120	4.5	450	14	US-10-081-816-57	Sequence 57, Appl
15	119	4.5	477	15	US-10-447-328-70	Sequence 70, Appl
16	118.5	4.5	2008	9	US-09-736-969A-2	Sequence 2, Appli
17	118.5	4.5	2008	9	US-09-736-969A-91	Sequence 91, Appl
18	118.5	4.5	2008	9	US-09-737-246-99	Sequence 99, Appl
19	118.5	4.5	2008	9	US-09-736-960-88	Sequence 88, Appl
20	118.5	4.5	2008	9	US-09-736-968A-105	Sequence 105, App
21	118.5	4.5	2008	10	US-09-978-244A-29	Sequence 29, Appl
22	115	4.3	367	14	US-10-081-816-13	Sequence 13, Appl
23	112.5	4.3	365	14	US-10-081-816-56	Sequence 56, Appl
24	112.5	4.3	408	15	US-10-447-328-50	Sequence 50, Appl
25	110.5	4.2	408	14	US-10-081-816-37	Sequence 37, Appl
26	110.5	4.2	2073	10	US-09-978-244A-10	Sequence 10, Appl
27	110	4.2	498	14	US-10-081-816-15	Sequence 15, Appl
28	108	4.1	381	14	US-10-081-816-31	Sequence 31, Appl
29	108	4.1	381	15	US-10-447-328-34	Sequence 34, Appl
30	107.5	4.1	346	15	US-10-282-122A-77576	Sequence 77576, A
31	107	4.0	404	14	US-10-081-816-52	Sequence 52, Appl
32	106.5	4.0	390	14	US-10-081-816-39	Sequence 39, Appl
33	106.5	4.0	390	15	US-10-447-328-58	Sequence 58, Appl
34	106.5	4.0	702	16	US-10-474-776-384	Sequence 384, App
35	106.5	4.0	702	17	US-10-472-928-4070	Sequence 4070, Ap
36	106.5	4.0	816	10	US-09-252-088-5	Sequence 5, Appli
37	106.5	4.0	816	15	US-10-340-792-5	Sequence 5, Appli
38	104	3.9	363	15	US-10-601-309-74	Sequence 74, Appl
39	103.5	3.9	960	14	US-10-214-390-4	Sequence 4, Appli
40	103	3.9	287	14	US-10-081-816-22	Sequence 22, Appl
41	103	3.9	1011	15	US-10-320-797-3324	Sequence 3324, Ap
42	102.5	3.9	373	14	US-10-081-816-4	Sequence 4, Appli
43	102.5	3.9	373	15	US-10-447-328-8	Sequence 8, Appli
44	102	3.9	320	10	US-09-510-332-160	Sequence 160, App
45	102	3.9	320	17	US-10-962-365-160	Sequence 160, App

ALIGNMENTS

RESULT 1
US-10-081-816-12
; Sequence 12, Application US/10081816
; Publication No. US20030045472A1
; GENERAL INFORMATION:
; APPLICANT: Axel, Richard
; APPLICANT: Scott, Kristin
; TITLE OF INVENTION: Chemosensory Gene Family Encoding Gustatory And Olfactory Recepto
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 0575/64019-A/JPW/ADM
; CURRENT APPLICATION NUMBER: US/10/081,816
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 60/271,319
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-081-816-12

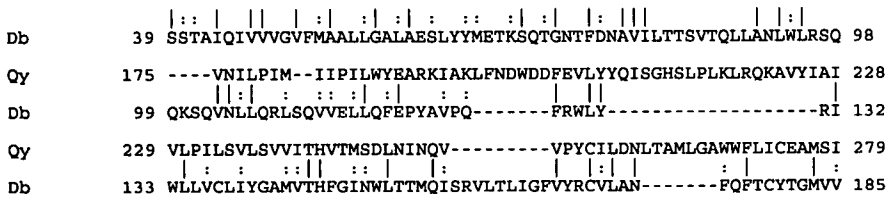
Query Match 100.0%; Score 2645; DB 14; Length 512;
Best Local Similarity 100.0%; Pred. No. 2.1e-252;
Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRPSGEKVVKGHGQGNSGHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRPSGEKVVKGHGQGNSGHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSI 60

Qy         61 GLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGP 120

Qy        121 ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPI 180

Qy        181 MIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 MIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVI 240

Qy        241 THVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAM 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 THVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAM 300

Qy        301 VADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLT 360

Qy        361 ITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPS 420

Qy        421 TINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLD 480

Qy        481 NDMSLMGASTLSTTTVGTTLPPPIMKLKGRKG 512
              ||||||||||||||||||||||||||||||||
Db        481 NDMSLMGASTLSTTTVGTTLPPPIMKLKGRKG 512


RESULT 2
US-10-081-816-24
; Sequence 24, Application US/10081816
; Publication No. US20030045472A1
; GENERAL INFORMATION:
;  APPLICANT: Axel, Richard
;  APPLICANT:  Scott, Kristin
;  TITLE OF INVENTION: Chemosensory Gene Family Encoding Gustatory And Olfactory Recepto
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 0575/64019-A/JPW/ADM
;  CURRENT APPLICATION NUMBER: US/10/081,816
;  CURRENT FILING DATE:  2002-02-22
;  PRIOR APPLICATION NUMBER: 60/271,319
;  PRIOR FILING DATE: 2001-02-23
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
;   LENGTH: 364
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-081-816-24

  Query Match             14.7%;  Score 390;  DB 14;  Length 364;
  Best Local Similarity   27.3%;  Pred. No. 1.4e-29;
  Matches  101;  Conservative   77;  Mismatches  152;  Indels   40;  Gaps    9;

Qy        110 IGVLPIVRHGP----ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRI-HIVRSLSGPF 164
              :||:|| |: |     |  :   |  ::::  :      |. |   |:  | |    |
Db          1 MGVMPIHRNPPEKNLPRTGYSWGSKQVMWAIFIYSCQTTIVVLVLRERVKKFVTSPDKRF 60

Qy        165 EEAVIAYLFLVNILPIMIIPIL-WYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKA 223
              :||:   :|:  :    ::|:  |    ::|   | | ::: ::: :|
Db         61 DEAIYNVIFISLLFTNFLLPVASWRHGPQVAIFKNMWTNYQYKFFKTTGS---------- 110

Qy        224 VYIAIVLPILSVLS---VVITHVTMSDLNINQVVPYCILDN----------LTAMLGAW 269
                || | |   |:    | : :    :|::|   | :  :          : |||  :
Db        111 ---PIVFPNLYPLTWSLCVFSWLLSIAINLSQ---YFLQPDFRLWYTFAYYPIIAMLNCF 164

Qy        270 ---WFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTF 326
                 |:: | |     :  |::  |  ::   ||  : :|| ||: || : :  | |    :
Db        165 CSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMY 224

Qy        327 VFMSLYLFFIITLSIYGLMSQ-LSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVR 385
              | :||   :: || :|: :  |   |::|| :   : :|||: ||:||||||  |
Db        225 GMYCLVIFFTTIIATYGSISEIIDHGATYKEVGLFVIVFYCMGLLYIICNEAHYASRKVG 284

Qy        386 TNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTY 445
              :|| ||| : |  :::  | |: | | |   ||  :|  |: ::|| |    :: | ||
Db        285 LDFQTKLLNINLTAVDAATQKEVEMLLVAINKNPPIMNLDGYANINRELITTNISFMATY 344

Qy        446 LVVLLQFQIS 455
              |||||||:|:
Db        345 LVVLLQFKIT 354


RESULT 3
US-10-447-328-10
; Sequence 10, Application US/10447328
; Publication No. US20040003419A1
; GENERAL INFORMATION:
;  APPLICANT: Carlson, John R.
;  APPLICANT:  Clyne, Peter J.
;  APPLICANT:  Warr, Coral G.
;  APPLICANT:  Yale University
;  TITLE OF INVENTION: No. US20040003419A1el Taste Receptors in Drosophila
;  FILE REFERENCE: 44574-5072
;  CURRENT APPLICATION NUMBER: US/10/447,328
;  CURRENT FILING DATE:  2003-05-29
;  PRIOR APPLICATION NUMBER: US/09/593,519
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/138,668
;  PRIOR FILING DATE: 1999-06-14
;  PRIOR APPLICATION NUMBER: US 60/181,704
;  PRIOR FILING DATE: 2000-02-10
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
;   LENGTH: 364
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-447-328-10

  Query Match             14.7%;  Score 390;  DB 15;  Length 364;
  Best Local Similarity   27.3%;  Pred. No. 1.4e-29;
  Matches  101;  Conservative   77;  Mismatches  152;  Indels   40;  Gaps    9;

Qy        110 IGVLPIVRHGP----ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRI-HIVRSLSGPF 164
              :||:|| |: |     |  :   |  ::::  :      |  |   |:  | |    |
Db          1 MGVMPIHRNPPEKNLPRTGYSWGSKQVMWAIFIYSCQTTIVVLVLRERVKKFVTSPDKRF 60

Qy        165 EEAVIAYLFLVNILPIMIIPIL-WYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKA 223
              :||:   :|:  :    ::|:  |    ::|   | | ::: ::: :|
Db         61 DEAIYNVIFISLLFTNFLLPVASWRHGPQVAIFKNMWTNYQYKFFKTTGS---------- 110

Qy        224 VYIAIVLPILSVLS---VVITHVTMSDLNINQVVPYCILDN----------LTAMLGAW 269
                || | |   |:    | : :    :|::|   | :  :          : |||  :
Db        111 ---PIVFPNLYPLTWSLCVFSWLLSIAINLSQ---YFLQPDFRLWYTFAYYPIIAMLNCF 164

Qy        270 ---WFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTF 326
                 |:: | |     :  |::  |  ::   ||  : :|| ||: || : :  | |    :
Db        165 CSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMY 224

Qy        327 VFMSLYLFFIITLSIYGLMSQ-LSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVR 385
              | :||   :: || :|: :  |   |::|| :   : :|||: ||:||||||  |
Db        225 GMYCLVIFFTTIIATYGSISEIIDHGATYKEVGLFVIVFYCMGLLYIICNEAHYASRKVG 284

Qy        386 TNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTY 445
              :|| ||| : |  :::  | |: | | |   ||  :|  |: ::|| |    :: | ||
Db        285 LDFQTKLLNINLTAVDAATQKEVEMLLVAINKNPPIMNLDGYANINRELITTNISFMATY 344

Qy        446 LVVLLQFQIS 455
              |||||||:|:
Db        345 LVVLLQFKIT 354


RESULT 4
US-10-081-816-67
; Sequence 67, Application US/10081816
; Publication No. US20030045472A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Axel, Richard
;  APPLICANT:  Scott, Kristin
;  TITLE OF INVENTION: Chemosensory Gene Family Encoding Gustatory And Olfactory Recepto
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 0575/64019-A/JPW/ADM
;  CURRENT APPLICATION NUMBER: US/10/081,816
;  CURRENT FILING DATE:  2002-02-22
;  PRIOR APPLICATION NUMBER: 60/271,319
;  PRIOR FILING DATE: 2001-02-23
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
;   LENGTH: 33
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-081-816-67

  Query Match             6.4%;  Score 169;  DB 14;  Length 33;
  Best Local Similarity   100.0%;  Pred. No. 3.3e-09;
  Matches   33;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        420 STINCGGFFDVNRTLFKGLLTTMVTYLVVLLQF 452
              |||||||||||||||||||||||||||||||||
Db          1 STINCGGFFDVNRTLFKGLLTTMVTYLVVLLQF 33


RESULT 5
US-10-447-328-80
; Sequence 80, Application US/10447328
; Publication No. US20040003419A1
; GENERAL INFORMATION:
;  APPLICANT: Carlson, John R.
;  APPLICANT:  Clyne, Peter J.
;  APPLICANT:  Warr, Coral G.
;  APPLICANT:  Yale University
;  TITLE OF INVENTION: No. US20040003419A1el Taste Receptors in Drosophila
;  FILE REFERENCE: 44574-5072
;  CURRENT APPLICATION NUMBER: US/10/447,328
;  CURRENT FILING DATE:  2003-05-29
;  PRIOR APPLICATION NUMBER: US/09/593,519
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/138,668
;  PRIOR FILING DATE: 1999-06-14
;  PRIOR APPLICATION NUMBER: US 60/181,704
;  PRIOR FILING DATE: 2000-02-10
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 80
;   LENGTH: 433
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-447-328-80

  Query Match             6.0%;  Score 160;  DB 15;  Length 433;
  Best Local Similarity   20.9%;  Pred. No. 1e-06;
  Matches   91;  Conservative   72;  Mismatches  139;  Indels  134;  Gaps   18;

Qy        102 PINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNR-IHIVRS- 159
              || :| | :       |   |     :  | :: |:   |:  ||   ||  : ||   :
Db         27 PIQFFTRTL-------HKRRRGIVILGYACYLISISLMVIYECYANIVALQKDIHKFHAE 79

Qy        160 ----LSGPF-EEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEV-LYYQISG- 212
                  : | : | :    :|: | | |::        |::|::::|  | |: |    ||
Db         80 DSSKVMGEYAERSWWVAMFVWNQLNILL------NFRRLARIYDDIADLEIDLNNASSGF 133

Qy        213 ----HSLPLKLR-QKAVYIAIVL-------------PILSVLSVVITHVTMSDLNINQV 253
                  |    : |   :| : |||             | |    : |:| : :   | : :
Db        134 VGQRHWWRFRFRLALSVGLWIVLLVGLTPRFTLVALGPYLHWTNKVLTEIILIMLQL-KC 192

Qy        254 VPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSK 313
               ||:          :  || | :    |:| ::    |:
Db        193 TEYCV----------FVLLIYELILRGRHIL-QQISVELE-----------------GN 223

Qy        314 LTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYI 373
              :||:   ||     ::|   ::   |:||::::|  |    : ||:  |:|   :  |
Db        224 QSRDSVQELC-----VALKRNQLLAGRIWGLVNEVSLYF---TLSLTLLFLYNELTILQI 275

Qy        374 CDEAHYASVN-------------VRTNFQKKLLMVELNWMNSDAQTEINMFLRA--TEMN 418
              : |   |||             : ||  | ::  :  : :     ||:||    :|
Db        276 VNWALIKSVNPNECCQYSKLVFKFKRNFTYKQVIFIIGRVGTCLLLSINIFLSCLYSEFC 335

Qy        419 PSTIN----------------------------------------CGGFFDVNRTLFKGL 438
              | |                                        ||| ||:|   | |:
Db        336 IQTYNSISRVLHQMYCLSAAEDYLILKMGLREYSLQMEHLKLIFTCGGLFDINLKFFGGM 395

Qy        439 LTTMVTYLVVLLQFQI 454
              : |:  |:::|:||:|
Db        396 VVTLFGYIIILVQFKI 411


RESULT 6
US-10-081-816-29
; Sequence 29, Application US/10081816
; Publication No. US20030045472A1
; GENERAL INFORMATION:
;  APPLICANT: Axel, Richard
;  APPLICANT:  Scott, Kristin
;  TITLE OF INVENTION: Chemosensory Gene Family Encoding Gustatory And Olfactory Recepto
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 0575/64019-A/JPW/ADM
;  CURRENT APPLICATION NUMBER: US/10/081,816
;  CURRENT FILING DATE:  2002-02-22
;  PRIOR APPLICATION NUMBER: 60/271,319
;  PRIOR FILING DATE: 2001-02-23
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 29
;   LENGTH: 369
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-081-816-29

  Query Match             5.1%;  Score 135.5;  DB 14;  Length 369;
  Best Local Similarity   20.5%;  Pred. No. 0.00021;
  Matches   87;  Conservative   65;  Mismatches  126;  Indels  147;  Gaps   20;

Qy        106 FLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFE 165
              :  ::|::|      |::||       :|| :  :|          | :|   |:  | :
Db         11 YFALLGLVP-WSESCAQSKFVQK----VYSAILIIL----------NAVHFGISIYFP-Q 54

Qy        166 EAVIAYLFLVNIL----PIMIIPILWYEARKIAKLFNDWDDF-----EVLYY------QI 210
              | :    :||::     |: : ::      | ::   :||:     |: |       ::
Db         55 SAELFLSLMVNVIVFVARIVCVTVI------ILQVMVHYDDYFRFCREMKYLGLRLQCEL 108

Qy        211 SGHSLPLKLRQKAVYIAI-------VLPIL-------------SVLSVVIT--------- 241
              |   || :  |  :|:        |||  :            |:||::|
Db        109 KIHVGRLKWQSYAKILALGIGFLVTVLPSIYVALSGSLLYFWSSLLSILIIRMQFVLVLL 168

Qy        242 -------HVTMSDLNINQVV------PYCILDNLTAMLGAWWFLICEAMS-ITAHLLAER 287
                    ||::  : :  |:        | ||      | :  ||:    | :  | |
Db        169 NVELLGHHVSLLGIRLQNVLECHLMGANCTLDGNANRLCSLEFLLALKQSHMQLHYLFTH 228

Qy        288 FQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQ 347
              |                                  | ::  |:|      ||   |::||
Db        229 FNDLF------------------------GWSILGTYV----VLFSDSTVNIYWTQQV 258

Qy        348 LSEGFGIKDIGLT----ITALWNIGLLFYICDE---------------AHYASVNVRTN 387
              | | :  | :  |    : : :|| |:|  | |                 : :|:   |:
Db        259 LVEVYEYKYLYATFSVFVPSFFNI-LVFCRCGEFCQRQSVLIGSYLRNLSCHPSIGRETS 317

Qy        388 FQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLV 447
              : | |||                |:  | |   ||  ||   : :|   :|   ||||:
Db        318 Y-KDLLM---------------EFILQVEQNVLAINAEGFMSTDNSLLMSILAAKVTYLI 361

Qy        448 VLLQF 452
```

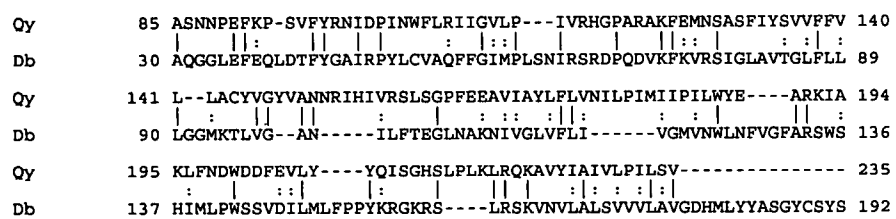
```
              ||:||
Db        362 VLMQF 366


RESULT 7
US-10-081-816-17
; Sequence 17, Application US/10081816
; Publication No. US20030045472A1
; GENERAL INFORMATION:
;  APPLICANT: Axel, Richard
;  APPLICANT:  Scott, Kristin
;  TITLE OF INVENTION: Chemosensory Gene Family Encoding Gustatory And Olfactory Recepto
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 0575/64019-A/JPW/ADM
;  CURRENT APPLICATION NUMBER: US/10/081,816
;  CURRENT FILING DATE:  2002-02-22
;  PRIOR APPLICATION NUMBER: 60/271,319
;  PRIOR FILING DATE: 2001-02-23
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
;   LENGTH: 736
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-081-816-17

  Query Match             5.0%;  Score 132.5;  DB 14;  Length 736;
  Best Local Similarity   21.0%;  Pred. No. 0.0011;
  Matches   77;  Conservative   66;  Mismatches  134;  Indels   89;  Gaps   16;

Qy        102 PINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNR-IHIVRS- 159
              || :| | :        |    |     :  | :: |:   |:  ||   ||  : ||   :
Db        427 PIQFFTRTL-------HKRRRGIVILGYACYLISISLMVIYECYANIVALQKDIHKFHAE 479

Qy        160 ----LSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEV-LYYQISG-- 212
                  : |  :: ::  :|: | | |::        |::|::::|  | |: |    ||
Db        480 DSSKVMGNTQKVLVVAMFVWNQLNILL------NFRRLARIYDDIADLEIDLNNASSGFV 533

Qy        213 ---HSLPLKLR-QKAVYIAIVL--------------PILSVLSVVITHVTMSDLNINQVV 254
                 |    : |   :| : |||               | |    : |:| : :  | : :
Db        534 GQRHWWRFRFRLALSVGLWIVLLVGLTPRFTLVALGPYLHWTNKVLTEIILIMLQL-KCT 592

Qy        255 PYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKL 314
              ||:          :  || | :    |:| ::    |:
Db        593 EYCV----------FVLLIYELILRGRHIL-QQISVELE-----------------GNQ 623

Qy        315 TRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYIC 374
              :||:   ||     ::|    ::   |:||::::|  |    : ||:  |:|   :  |
Db        624 SRDSVQELC-----VALKRNQLLAGRIWGLVNEVSLYF---TLSLTLLFLYNELTILQIV 675

Qy        375 DEAHYASVNVR--TNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNR 432
              : |   |||      : :  |::::        | |     :                ||| ||:|
Db        676 NWALIKSVNPNECCQYTEDYLILKMGLREYSLQMEHLKLI---------FTCGGLFDINL 726

Qy        433 TLFKGL 438
                | |:
Db        727 KFFGGV 732


RESULT 8
US-10-081-816-30
; Sequence 30, Application US/10081816
; Publication No. US20030045472A1
; GENERAL INFORMATION:
;  APPLICANT: Axel, Richard
;  APPLICANT:  Scott, Kristin
;  TITLE OF INVENTION: Chemosensory Gene Family Encoding Gustatory And Olfactory Recepto
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 0575/64019-A/JPW/ADM
;  CURRENT APPLICATION NUMBER: US/10/081,816
;  CURRENT FILING DATE:  2002-02-22
;  PRIOR APPLICATION NUMBER: 60/271,319
;  PRIOR FILING DATE: 2001-02-23
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
;   LENGTH: 372
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-081-816-30

  Query Match             4.9%;  Score 130.5;  DB 14;  Length 372;
  Best Local Similarity   21.7%;  Pred. No. 0.00067;
  Matches   80;  Conservative   64;  Mismatches  135;  Indels   89;  Gaps   17;

Qy        129 SASFIYSVVFFVLLACYVGYVANNRIHI-VRSLSG-PFEEAVI---------AYLFL--- 174
              |:: |  ||  | :|  :| :| :  ::  :| :|  |: |||         | |:|
Db         39 SSTAIQIVVVGVFMAALLGALAESLYYMETKSQTGNTFDNAVILTTSVTQLLANLWLRSQ 98

Qy        175 ----VNILPIM--IIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAI 228
                  ||:|  :  :: :| :|   : :       |  ||                   |
Db         99 QKSQVNLLQRLSQVVELLQFEPYAVPQ-------FRWLY------------------RI 132

Qy        229 VLPILSVLSVVITHVTMSDLNINQV---------VPYCILDNLTAMLGAWWFLICEAMSI 279
               | :  :   ::||  :: |   |:         |  |:| |       : |     | :
Db        133 WLLVCLIYGAMVTHFGINWLTTMQISRVLTLIGFVYRCVLAN-------FQFTCYTGMVV 185

Qy        280 TAHLLAERFQKALKHI---------GPAAMVADY-RVLWLRLSKLTRDTGNALCYTFVF- 328
                 | :   | |:|:          | |  :  :  :| |   :|     |  : : |::
Db        186 ILKKLLQVQVKQLEHLVSTTTISMAGVAGCLRTHDEILLLGQRELIAVYGGVILFLFIYQ 245

Qy        329 --MSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVRT 386
                 : :|:|  |           |||   :  | :   |   :|||:         :: :
Db        246 VMQCILIFYISNL----------EGFHSSN-DLVLIFCWLAPMLFYLILPLVVNDIHNQA 294

Qy        387 NFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNR-TLFKGLLTTMVTY 445
              |   |:|  ::    :     |  ||         :   ||| ::: |||| | | : ||
Db        295 NKTAKML-TKVPRTGTGLDRMIEKFLLKNLRQKPILTAYGFFALDKSTLFK-LFTAIFTY 352

Qy        446 LVVLLQFQ 453
              :|:|:||:
Db        353 MVILVQFK 360


RESULT 9
US-10-447-328-32
; Sequence 32, Application US/10447328
; Publication No. US20040003419A1
; GENERAL INFORMATION:
;  APPLICANT: Carlson, John R.
;  APPLICANT:  Clyne, Peter J.
;  APPLICANT:  Warr, Coral G.
;  APPLICANT:  Yale University
;  TITLE OF INVENTION: No. US20040003419A1el Taste Receptors in Drosophila
;  FILE REFERENCE: 44574-5072
;  CURRENT APPLICATION NUMBER: US/10/447,328
;  CURRENT FILING DATE:  2003-05-29
;  PRIOR APPLICATION NUMBER: US/09/593,519
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/138,668
;  PRIOR FILING DATE: 1999-06-14
;  PRIOR APPLICATION NUMBER: US 60/181,704
;  PRIOR FILING DATE: 2000-02-10
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
;   LENGTH: 372
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-447-328-32

  Query Match             4.9%;  Score 130.5;  DB 15;  Length 372;
  Best Local Similarity   21.7%;  Pred. No. 0.00067;
  Matches   80;  Conservative   64;  Mismatches  135;  Indels   89;  Gaps   17;

Qy        129 SASFIYSVVFFVLLACYVGYVANNRIHI-VRSLSG-PFEEAVI---------AYLFL--- 174
```



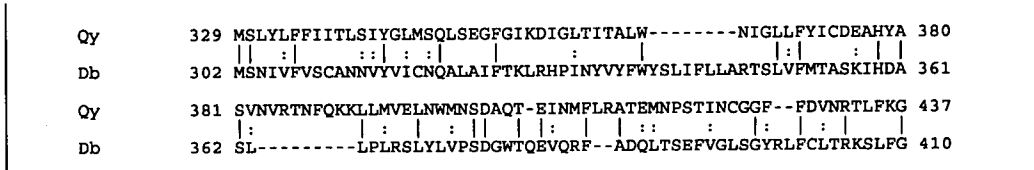
```
Db         39 SSTAIQIVVVGVFMAALLGALAESLYYMETKSQTGNTFDNAVILTTSVTQLLANLWLRSQ 98

Qy        175 ----VNILPIM--IIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAI 228

Db         99 QKSQVNLLQRLSQVVELLQFEPYAVPQ-------FRWLY------------------RI 132

Qy        229 VLPILSVLSVVITHVTMSDLNINQV---------VPYCILDNLTAMLGAWWFLICEAMSI 279

Db        133 WLLVCLIYGAMVTHFGINWLTTMQISRVLTLIGFVYRCVLAN-------FQFTCYTGMVV 185

Qy        280 TAHLLAERFQKALKHI---------GPAAMVADY-RVLWLRLSKLTRDTGNALCYTFVF- 328

Db        186 ILKKLLQVQVKQLEHLVSTTTISMAGVAGCLRTHDEILLLGQRELIAVYGGVILPLFIYQ 245

Qy        329 --MSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVRT 386

Db        246 VMQCILIFYISNL----------EGFHSSN-DLVLIFCWLAPMLFYLILPLVVNDIHNQA 294

Qy        387 NFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNR-TLFKGLLTTMVTY 445

Db        295 NKTAKML-TKVPRTGTGLDRMIEKFLLKNLRQKPILTAYGFFALDKSTLFK-LFTAIFTY 352

Qy        446 LVVLLQFQ 453

Db        353 MVILVQFK 360
```

RESULT 10
US-10-081-816-11
; Sequence 11, Application US/10081816
; Publication No. US20030045472A1
; GENERAL INFORMATION:
; APPLICANT: Axel, Richard
; APPLICANT: Scott, Kristin
; TITLE OF INVENTION: Chemosensory Gene Family Encoding Gustatory And Olfactory Receptc
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 0575/64019-A/JPW/ADM
; CURRENT APPLICATION NUMBER: US/10/081,816
; CURRENT FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: 60/271,319
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-081-816-11

Query Match 4.9%; Score 129; DB 14; Length 436;
Best Local Similarity 21.2%; Pred. No. 0.0012;
Matches 94; Conservative 70; Mismatches 172; Indels 108; Gaps 22;



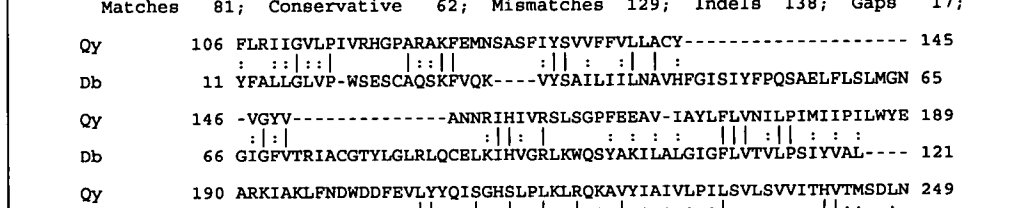
```
Qy         85 ASNNPEFKP-SVFYRNIDPINWFLRIIGVLP---IVRHGPARAKFEMNSASFIYSVVFFV 140

Db         30 AQGGLEFEQLDTFYGAIRPYLCVAQFFGIMPLSNIRSRDPQDVKFKVRSIGLAVTGLFLL 89

Qy        141 L--LACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYE----ARKIA 194

Db         90 LGGMKTLVG--AN-----ILFTEGLNAKNIVGLVFLI------VGMVNWLNFVGFARSWS 136

Qy        195 KLFNDWDDFEVLY----YQISGHSLPLKLRQKAVYIAIVLPILSV--------------- 235

Db        137 HIMLPWSSVDILMLFPPYKRGKRS----LRSKVNVLALSVVVLAVGDHMLYYASGYCSYS 192

Qy        236 LSVVITHVTMSDL-----------NINQVVPYCILDNLTAM-LGAW--------WFLICE 275

Db        193 MHILQCHTNHSRITFGLYLEKEFSDIMFIMPF----NIFSMCYGFWLNGAFTFLWNFMDI 248

Qy        276 AMSITAHLLAERFQKALKHIG-------PAAMVADYRVLWLRLSKLTRDTGNALCYTFVF 328

Db        249 FIVMTSIGLAQRFQQFAARVGALEGRHVPEALWYDIRRDHIRLCELA-------SLVEAS 301
```



```
Qy        329 MSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALW--------NIGLLFYICDEAHYA 380

Db        302 MSNIVFVSCANNVYVICNQALAIFTKLRHPINYVYFWYSLIFLLARTSLVFMTASKIHDA 361

Qy        381 SVNVRTNFQKKLLMVELNWMNSDAQT-EINMFLRATEMNPSTINCGGF--FDVNRTLFKG 437

Db        362 SL---------LPLRSLYLVPSDGWTQEVQRF--ADQLTSEFVGLSGYRLFCLTRKSLFG 410

Qy        438 LLTTMVTYLVVLLQFQISIPTDKG 461

Db        411 MLATLVTYELMLL--QIDAKSHKG 432
```

RESULT 11
US-10-447-328-30
; Sequence 30, Application US/10447328
; Publication No. US20040003419A1
; GENERAL INFORMATION:
; APPLICANT: Carlson, John R.
; APPLICANT: Clyne, Peter J.
; APPLICANT: Warr, Coral G.
; APPLICANT: Yale University
; TITLE OF INVENTION: No. US20040003419A1el Taste Receptors in Drosophila
; FILE REFERENCE: 44574-5072
; CURRENT APPLICATION NUMBER: US/10/447,328
; CURRENT FILING DATE: 2003-05-29
; PRIOR APPLICATION NUMBER: US/09/593,519
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/138,668
; PRIOR FILING DATE: 1999-06-14
; PRIOR APPLICATION NUMBER: US 60/181,704
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-10-447-328-30

Query Match 4.8%; Score 127; DB 15; Length 348;
Best Local Similarity 19.8%; Pred. No. 0.0013;
Matches 81; Conservative 62; Mismatches 129; Indels 138; Gaps 17;



```
Qy        106 FLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACY---------- 145

Db         11 YFALLGLVP-WSESCAQSKFVQK----VYSAILIILNAVHFGISIYFPQSAELFLSLMGN 65

Qy        146 -VGYV-------------ANNRIHIVRSLSGPFEEAV-IAYLFLVNILPIMIIPILWYE 189

Db         66 GIGFVTRIACGTYLGLRLQCELKIHVGRLKWQSYAKILALGIGFLVTVLPSIYVAL---- 121

Qy        190 ARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTMSDLN 249

Db        122 ------------SGSLLYFWSSLLSI-LIIRMQFVLVLLNVELLG------HHVSLLGIR 162

Qy        250 INQVV------PYCILDNLTAMLGAWWFLICEAMS-ITAHLLAERFQKALKHIGPAAMVA 302

Db        163 LQNVLECHLMGANCTLDGNANRLCSLEFLLALKQSHMQLHYLFTHFNDLF---------- 212

Qy        303 DYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLT-- 360

Db        213 ---------------GWSILGTYV----VLFSDSTVNIYWTQQVLVEVYEYKYLYATFS 252

Qy        361 --ITALWNIGLLFYICDE---------------AHYASVNVRTNFQKKLLMVELNWMNS 402

Db        253 VFVPSFFNI-LVFCRCGEFCQRQSVLIGSYLRNLSCHPSIGRETSY-KDLLM-------- 302

Qy        403 DAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLLQF 452

Db        303 -------EFILQVEQNVLAINAEGFMSTDNSLLMSILAAKVTYLIVLMQF 345
```

```
RESULT 12
US-10-081-816-1
; Sequence 1, Application US/10081816
; Publication No. US20030045472A1
; GENERAL INFORMATION:
;  APPLICANT: Axel, Richard
;  APPLICANT:  Scott, Kristin
;  TITLE OF INVENTION: Chemosensory Gene Family Encoding Gustatory And Olfactory Recepto
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 0575/64019-A/JPW/ADM
;  CURRENT APPLICATION NUMBER: US/10/081,816
;  CURRENT FILING DATE:  2002-02-22
;  PRIOR APPLICATION NUMBER: 60/271,319
;  PRIOR FILING DATE: 2001-02-23
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 410
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-081-816-1

  Query Match             4.6%;  Score 122;  DB 14;  Length 410;
  Best Local Similarity   22.2%;  Pred. No. 0.0053;
  Matches   70;  Conservative   53;  Mismatches  104;  Indels   88;  Gaps   18;

Qy        185 ILW-YEARKI----AKLFNDWDDPEVLYYQISGHSLPL---KLRQKAVYIAIVLPILSVL 236
              ||| |   ::    ||| :  | | :  | || : |||    ||   :: |  | :   |
Db        149 ILWGYAVSQLLILGAKLLSRGDRFPI--YWIS-YLLPLLVCGLRYFQIFNATQL-VRQRL 204

Qy        237 SVVITHVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIG 296
              |::  | :  | ::|  |                      |:        :  : |:  :
Db        205 DVLL--VALQQLQLHQKGP--------------------AVDTVLEEQEDLEEAAMDRLI 242

Qy        297 PAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSLYL-----FFIITLSIYGLMSQLSEG 351
                :|  |: :|  :: | |       ||    :|: :     |  || : | :    :
Db        243 AVRLV--YQRVWALVALLNR------CYG---LSMLMQVGNDFLAITSNCYWMFLNFRQS 291

Qy        352 ----FGIKDIGLTITALW------NIGLLFYICDEA-----------HYASVNVRTNFQK 390
                | |  : :  : :|       |: :|  :||            |  ||::|
Db        292 AASPFDI--LQIVASGVWSAPHLGNVLVLSLLCDRTAQCASRLALCLHQVSVDLRNESHN 349

Qy        391 KLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLL 450
              |:             |: :: |    ::    :  |||:|: ||   ::    |||::|:
Db        350 ALI------------TQFSLQLLHQRLH---FSAAGFFNVDCTLLYTIVGATTTYLIILI 394

Qy        451 QFQISIPTDKGDSEG 465
              || :|  |   || |
Db        395 QFHMSESTIGSDSNG 409


RESULT 13
US-10-447-328-2
; Sequence 2, Application US/10447328
; Publication No. US20040003419A1
; GENERAL INFORMATION:
;  APPLICANT: Carlson, John R.
;  APPLICANT:  Clyne, Peter J.
;  APPLICANT:  Warr, Coral G.
;  APPLICANT:  Yale University
;  TITLE OF INVENTION: No. US20040003419A1el Taste Receptors in Drosophila
;  FILE REFERENCE: 44574-5072
;  CURRENT APPLICATION NUMBER: US/10/447,328
;  CURRENT FILING DATE:  2003-05-29
;  PRIOR APPLICATION NUMBER: US/09/593,519
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/138,668
;  PRIOR FILING DATE: 1999-06-14
;  PRIOR APPLICATION NUMBER: US 60/181,704
;  PRIOR FILING DATE: 2000-02-10
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 414
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-447-328-2

  Query Match             4.6%;  Score 122;  DB 15;  Length 414;
  Best Local Similarity   22.2%;  Pred. No. 0.0054;
  Matches   70;  Conservative   53;  Mismatches  104;  Indels   88;  Gaps   18;

Qy        185 ILW-YEARKI----AKLFNDWDDFEVLYYQISGHSLPL---KLRQKAVYIAIVLPILSVL 236
              ||| |   ::    ||| :  | | :  | || : |||    ||   :: |  | :   |
Db        153 ILWGYAVSQLLILGAKLLSRGDRFPI--YWIS-YLLPLLVCGLRYFQIFNATQL-VRQRL 208

Qy        237 SVVITHVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIG 296
              |::  | :  | ::|  |                      |:        :  : |:  :
Db        209 DVLL--VALQQLQLHQKGP--------------------AVDTVLEEQEDLEEAAMDRLI 246

Qy        297 PAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSLYL-----FFIITLSIYGLMSQLSEG 351
                :|  |: :|  :: | |       ||    :|: :     |  || : | :    :
Db        247 AVRLV--YQRVWALVALLNR------CYG---LSMLMQVGNDFLAITSNCYWMFLNFRQS 295

Qy        352 ----FGIKDIGLTITALW------NIGLLFYICDEA-----------HYASVNVRTNFQK 390
                | |  : :  : :|       |: :|  :||            |  ||::|
Db        296 AASPFDI--LQIVASGVWSAPHLGNVLVLSLLCDRTAQCASRLALCLHQVSVDLRNESHN 353

Qy        391 KLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLL 450
              |:             |: :: |    ::    :  |||:|: ||   ::    |||::|:
Db        354 ALI------------TQFSLQLLHQRLH---FSAAGFFNVDCTLLYTIVGATTTYLIILI 398

Qy        451 QFQISIPTDKGDSEG 465
              || :|  |   || |
Db        399 QFHMSESTIGSDSNG 413


RESULT 14
US-10-081-816-57
; Sequence 57, Application US/10081816
; Publication No. US20030045472A1
; GENERAL INFORMATION:
;  APPLICANT: Axel, Richard
;  APPLICANT:  Scott, Kristin
;  TITLE OF INVENTION: Chemosensory Gene Family Encoding Gustatory And Olfactory Recepto
;  TITLE OF INVENTION:  Uses Thereof
;  FILE REFERENCE: 0575/64019-A/JPW/ADM
;  CURRENT APPLICATION NUMBER: US/10/081,816
;  CURRENT FILING DATE:  2002-02-22
;  PRIOR APPLICATION NUMBER: 60/271,319
;  PRIOR FILING DATE: 2001-02-23
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
;   LENGTH: 450
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-081-816-57

  Query Match             4.5%;  Score 120;  DB 14;  Length 450;
  Best Local Similarity   19.0%;  Pred. No. 0.0096;
  Matches   91;  Conservative   76;  Mismatches  149;  Indels  162;  Gaps   26;

Qy         71 EAPPLDRKKSSDSTASNN-------PEF-KPSVFYRNIDPINWFLRIIGVLPIV---RHG 119
              | ||   |   ||    |       | :    :|:| : |  :  : : ::|:|
Db         28 ETPP--PKFVEDSNLEFNVLASEKLPNYTNLDLFHRAVFPFMFLAQCVAIMPLVGIRESN 85

Qy        120 PARAKFEMNSASFIYSVVFFV-----LLACY-----VGYVANNRIHIVRSLSGPFEEAVI 169
              | | :|   |     :::| :      |: :     :|  | | : :|      |   :
Db         86 PRRVRFAYKSIPMFVTLIFMIATSILFLSMFTHLLKIGITAKNFVGLVF-----FGCVLS 140

Qy        170 AYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEVLY----YQISGHSLP--------- 216
              ||:  : :            |:|   :   |   |: :     |:|  :|
Db        141 AYVVFIRL-----------AKKWPAVVRIWTRTEIPFTKPPYEIPKRNLSRRVQLAALA 188
```

```
Qy        217 ---LKLRQKAVY-IAIVLPILSVLSVVITHVTMSDLN---------INQVVPY------C 257
Db        189 IIGLSLGEHALYQVSAILSYTRRIQMCANITTVPSFNNYMQTNYDYVFQLLPYSPIIAVL 248

Qy        258 ILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRD 317
Db        249 ILATCTFV---WNYMDLFIMMISKG-LSYRFEQ----------------ITTRIRKLEHE 288

Qy        318 TGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWN-IGLLFYICDE 376
Db        289 ---EVCES--------VFIQIREHYVKMCELLE---FVDSAMSSLILLSCVNNLYFVC-- 332

Qy        377 AHYASVNVRTNFQKKLLMVELNW---------------------------MNSDAQTEIN 409
Db        333 --YQLLNV---FNK------LRWPINYIYFWYSLLYLIGRTAFVFLTAADINEESKRGLG 381

Qy        410 MFLRAT------------EMNPSTINCGG--FFDVNRTLFKGLLTTMVTYLVVLLQF 452
Db        382 VLRRVSSRSWCVEVERLIFQMTTQTVALSGKKFYFLTRRLLFGMAGTIVTYELVLLQF 439


RESULT 15
US-10-447-328-70
; Sequence 70, Application US/10447328
; Publication No. US20040003419A1
; GENERAL INFORMATION:
;  APPLICANT: Carlson, John R.
;  APPLICANT:  Clyne, Peter J.
;  APPLICANT:  Warr, Coral G.
;  APPLICANT:  Yale University
;  TITLE OF INVENTION: No. US20040003419A1el Taste Receptors in Drosophila
;  FILE REFERENCE: 44574-5072
;  CURRENT APPLICATION NUMBER: US/10/447,328
;  CURRENT FILING DATE:  2003-05-29
;  PRIOR APPLICATION NUMBER: US/09/593,519
;  PRIOR FILING DATE: 2000-06-14
;  PRIOR APPLICATION NUMBER: US 60/138,668
;  PRIOR FILING DATE: 1999-06-14
;  PRIOR APPLICATION NUMBER: US 60/181,704
;  PRIOR FILING DATE: 2000-02-10
;  NUMBER OF SEQ ID NOS: 94
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 70
;   LENGTH: 477
;   TYPE: PRT
;   ORGANISM: Drosophila melanogaster
US-10-447-328-70

  Query Match             4.5%;  Score 119;  DB 15;  Length 477;
  Best Local Similarity   19.9%;  Pred. No. 0.013;
  Matches  101;  Conservative   73;  Mismatches  171;  Indels  162;  Gaps   22;

Qy        108 RIIGVLP-IVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIV------RSL 160
Db         23 KIAGILPQDLEKFRSRNLLEKSRNGMIY------MLSTLILYVVLYNILIYSFGEEDRSL 76

Qy        161 SGPFEEAVIAY---LFLVNILPIMIIPILWYEARKIAKLFNDWDDFEVLYYQI------- 210
Db         77 KA--SQSTLTFVIGLFLTYIGLIMMV------SDQLTALRNQGRIGE-LYERIRLVDERL 127

Qy        211 --SGHSLPLKLRQKAVYIAIVLPILSVLSVVI-THVTMSD-------LNINQVVPYCI-- 258
Db        128 YKEGCVMDNSTIGRRIRIMLIMTVIFELSILVSTYVKLVDYSQWMSLLWIVSAIPTFINT 187

Qy        259 LDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLR-------- 310
Db        188 LDKI-------WF------AVSLYALKERFEAINATLEELVDTHEKHKLWLRGNQEVPPP 234

Qy        311 ------------LSKLTRDTG----------------NALC------------------Y 324
Db        235 LDSSQPPQYDSNLEYLYKELGGMDIGSIGKSSVEEKLNNLCQVHDEICEIGKALNELWSY 294

Qy        325 TFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNV 384
Db        295 PILSLMAYGFLIFTAQLYFL--------------YCATQYQSIPSLFRSAKNPFITVIVL 340

Qy        385 RTNFQKKLLMVELNWMNSDAQT------------------EINMFLRATEMNPST-INC 424
Db        341 SYTSGKCVYLIYLSWKTSQASKRTGISLHKCGVVADDNLLYEIVNHLSLKLLNHSVDFSA 400

Qy        425 GGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLDNDMS 484
Db        401 CGFFTLDMETLYGVSGGITSYLIILIQFNLAAQQAKEAIQTFN----------SLNDTAG 450

Qy        485 LMGAST----LSTTT---VGTTLPPPI 504
Db        451 LVGAATDMDNISSTLRDFVTTTMTPAV 477


Search completed: May 23, 2005, 19:29:12
Job time : 144 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 23, 2005, 18:59:59 ; Search time 41 Seconds
(without alignments)
1201.536 Million cell updates/sec

Title: US-10-081-816-12
Perfect score: 2645
Sequence: 1 MRPSGEKVVKGHGQGNSGHS.....TTTVGTTLPPPIMKLKGRKG 512

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	119.5	4.5	439	2	C22845	NADH2 dehydrogenas
2	118.5	4.5	789	2	T29461	hypothetical prote
3	112	4.2	443	2	F30010	NADH2 dehydrogenas
4	111.5	4.2	395	2	I64248	hypothetical prote
5	107.5	4.1	346	2	H82509	hypothetical prote
6	107.5	4.1	583	2	AI1510	C-terminal domain
7	106.5	4.0	698	2	D95228	hypothetical prote
8	106.5	4.0	702	2	H98092	hypothetical prote
9	105.5	4.0	653	2	B71640	NADH2 dehydrogenas
10	105	4.0	453	2	H71732	cytochrome D ubiqu
11	105	4.0	468	2	A75145	protein translocas
12	105	4.0	654	2	E64245	periplasmic phosph
13	105	4.0	766	2	B75059	probable transmemb
14	104.5	4.0	896	2	H86762	ABC transporter pe
15	104	3.9	369	1	G70007	conserved hypothet
16	103.5	3.9	317	2	T21047	hypothetical prote
17	103.5	3.9	459	2	T17358	NADH2 dehydrogenas
18	103.5	3.9	515	2	AD2349	hypothetical prote
19	103.5	3.9	656	2	T31884	hypothetical prote
20	102.5	3.9	598	2	D71127	hypothetical prote
21	102.5	3.9	725	2	A90255	hypothetical prote
22	102	3.9	468	2	E71184	probable preprotei
23	102	3.9	925	2	A72096	ct234 hypothetical
24	102	3.9	925	2	E81573	conserved hypothet
25	102	3.9	925	2	E86527	CT234 hypothetical
26	102	3.9	1036	2	D70117	acriflavine resist
27	101.5	3.8	437	2	S34959	NADH2 dehydrogenas
28	101.5	3.8	437	2	S73284	hypothetical prote
29	101.5	3.8	461	1	QXXL4M	NADH2 dehydrogenas
30	101.5	3.8	488	2	T20124	hypothetical prote
31	101.5	3.8	511	2	F84968	virulence factor m
32	101.5	3.8	891	2	A97718	hypothetical prote
33	101	3.8	459	2	A59154	NADH2 dehydrogenas
34	101	3.8	651	2	F90536	transport protein
35	100.5	3.8	467	2	AG0546	probable terminal
36	100.5	3.8	580	2	A83874	carbon starvation-
37	100.5	3.8	1483	2	S42839	T16G12.5 protein -
38	100	3.8	379	2	PC4180	thiazide-sensitive
39	100	3.8	429	1	C41903	arsenical pump mem
40	100	3.8	653	2	T14241	NADH2 dehydrogenas
41	99.5	3.8	244	2	B64014	hypothetical prote
42	99.5	3.8	301	2	G96944	probable permease
43	99.5	3.8	583	2	AH1151	glycerophosphoryl
44	99	3.7	348	2	T21627	hypothetical prote
45	99	3.7	360	2	B46647	melanocortin recep

ALIGNMENTS

RESULT 1
C22845
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Trypanosoma brucei mitochondrion
C;Species: mitochondrion Trypanosoma brucei
C;Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 09-Jul-2004
C;Accession: C22845
R;Hensgens, L.A.M.; Brakenhoff, J.; De Vries, B.F.; Sloof, P.; Tromp, M.C.; Van Boom, J.H
Nucleic Acids Res. 12, 7327-7344, 1984
A;Title: The sequence of the gene for cytochrome c oxidase subunit I, a frameshift conta
ondrial maxi-circle DNA.
A;Reference number: A93537; MUID:85037915; PMID:6093040
A;Accession: C22845
A;Molecule type: DNA
A;Residues: 1-439 <HEN>
A;Cross-references: UNIPROT:Q33575
C;Genetics:
A;Genome: mitochondrion
A;Genetic code: SGC6
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 4
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 4.5%; Score 119.5; DB 2; Length 439;
Best Local Similarity 16.4%; Pred. No. 0.055;
Matches 71; Conservative 82; Mismatches 127; Indels 153; Gaps 17;

```
Qy      105 WFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVPFVLLACYVGYVANNRIHIVRSLSGPF 164
            ||:  :|::                  :| ::| :   |    |: |:   :
Db       51 WFVFFMGII------------------MYILIFLLSKKC----VSYNKYFYI------- 80

Qy      165 EEAVIAYLFL-VNILPIMIIPILWYEARKIAKLFNDWDD-------FEVLYYQISGHSLP 216
               |: |::: :|:: |:|:                  ||       || |:: |   ||
Db       81 ---VMIYMYIYINVVLIIIL-----------------DDFMCFMIAFESLFFPICLVSLF 120

Qy      217 LKLRQKAV----YIAIVLPILSVLSVVITHVTMSD---LNINQVVPYCILDNLTAMLGAW 269
                 : :    |: |   : ||: ::|  : :|    :|:   :  |  |:| : :  |
Db      121 FNFNNRFIFAIFYLIIFSSVSSVVCIIICIIVISHFNIINLQAFIDVCYFDSLYSAIFIW 180

Qy      270 WFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWL-------------RLSKLTR 316
              |                |:| | |          :||               |: :
Db      181 ILLF--------------IMFAIKPIWPFH-------VWLPEMHVEVNTEMSVLLASIVL 219

Qy      317 DTGNALCYTFVFMSL------YLFFIITLSIYGLM------------SQLSEGPGIKDIG 358
              |    | |:|::       :| || :: : ||:                ::   : | |
Db      220 KIGFFGVYKFLFIAFNTISIWFLGFIDSVIVLGLVFIAMSLIFLSDYKKIIANWSIIHTG 279

Qy      359 LTITALWNIGLLF----YICDEAHYASVN--------VRTNFQKKLLMVELNWMNSDAQT 406
            : :  ||:  :||      :|: ||  | :         :  |:  :: :: :::      :
Db      280 IGLILLWHNDILFVGLLILCNLAHILSSSFMFIVIGYMYDNYGVRIFLLLISFFGISIWS 339

Qy      407 E------------------INMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLV 447
                             :::|:    :: | |    |: :  |:|   :     |:
Db      340 SLFLCLFLFNIDFPFMLLFYVDIFILYGLISISFIYIISFYIITLTIFLSSI-----YIY 394
```

```
Qy        448 VLLQFQISIPTDK 460
              : | |    :  ||
Db        395 MCLSFYSFVWLDK 407


RESULT 2
T29461
hypothetical protein T05B11.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T29461
R;Geisel, C.; Bradshaw, H.
submitted to the EMBL Data Library, March 1996
A;Description: The sequence of C. elegans cosmid T05B11.
A;Reference number: Z20621
A;Accession: T29461
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-789 <GEI>
A;Cross-references: UNIPROT:Q22211; EMBL:U53151; PIDN:AAB37067.1; GSPDB:GN00023; CESP:T0
A;Experimental source: strain Bristol N2; clone T05B11
C;Genetics:
A;Gene: CESP:T05B11.2
A;Map position: 5
A;Introns: 29/3; 123/2; 229/2; 344/1; 362/1; 418/3; 512/2; 618/2; 707/2

  Query Match             4.5%;  Score 118.5;  DB 2;  Length 789;
  Best Local Similarity   20.8%;  Pred. No. 0.14;
  Matches  100;  Conservative   65;  Mismatches  170;  Indels  145;  Gaps   20;

Qy         37 FLNAKPL--------NSANAQAYLYGVRKYSIGLAERLDADYEAPPLDRKKSSDSTASNN 88
              | |: ||        | |    |:: |      | | |             ||::    ||
Db        230 FFNSNPLLYSIILGFNLAAIIFYIPFVLIPPCTLQEHL-----------KSTNCILINN 277

Qy         89 PEFK----PSVF--YR----NIDPINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVF 138
               ||:    | |:  ||     :  |   : ||  ||| ::  |   | : :  |:  ::
Db        278 DEFECTKDPIVYQTYRIMIDRLQKIETRICIIASLPITQNTFAACTFIIPNFGFLLIIIK 337

Qy        139 FVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFN 198
              |:|       ||    ::  :|                    :|:  ::  | ::|
Db        338 KVIL-------ANELDKLMVKMSN-------------------LPVFSFDDDKSSELLG 370

Qy        199 DWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTMSDLNINQVVPYCI 258
               :  |     ::|      | :  :::  :|     |: |:|| : :| |      |:
Db        371 PFRHF----LKVSLLDCSAKTKCCSIFTCLV----GVIIVIITFMRISVLMQTAGAPFSF 422

Qy        259 LDNLTAMLG--AWWFLICE--AMSITAHLLAERFQKALKHIGPAAMVA-----DYRVLWL 309
                :  | | :  :|       | : || :| | |  :     ||      ::|:| :
Db        423 SWGEASFFGYPAIFGFVCSLCLFGWTKNGLASKFCKRLVRVRMLRQVANPKLDNFRILHI 482

Qy        310 RLSKLTRDTGNALCYTF----VFMSLYLFFIITLSIY--GLMSQLSEGFGIKDIGLTITA 363
                    |||::       ||  ::  |   ||  |:   |      |    | |
Db        483 L----------ALCFSIPWFVAIMSWIIYNFINGKIYYGGIEQNLLSRIFILVTNLYIWY 532

Qy        364 LWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQ------------------ 405
              :  | |  ||                |     |  |:|: | : |
Db        533 ISTICLAIYI--------------FMSSALNREVNYFNEELQKAKEEKTLKNIGVLEKFD 578

Qy        406 ------TEINMFLRATEMNPSTINCGGFFDVNRTLFKGLL----------TTMVTYLVVL 449
                    |: :|       | |  : |||  :   |  ||:          || | | ::|
Db        579 FRQNEILEMTLF-----ANGSLSSLGGFAPL--FLLWGLVNGIYLTSFFNTTPVLYCIIL 631


RESULT 3
F30010
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Leishmania tarentolae mitochondr
C;Species: mitochondrion Leishmania tarentolae
C;Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 03-Jun-2002
C;Accession: F30010
R;de la Cruz, V.F.; Neckelmann, N.; Simpson, L.
J. Biol. Chem. 259, 15136-15147, 1984
A;Title: Sequences of six genes and several open reading frames in the kinetoplast maxid
A;Reference number: A22848; MUID:85079995; PMID:6096360
A;Accession: F30010
A;Molecule type: DNA
A;Residues: 1-443 <DEL>
A;Cross-references: GB:M10126
C;Genetics:
A;Genome: mitochondrion
A;Genetic code: SGC6
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 4
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation; c

  Query Match             4.2%;  Score 112;  DB 2;  Length 443;
  Best Local Similarity   20.2%;  Pred. No. 0.23;
  Matches   82;  Conservative   65;  Mismatches  131;  Indels  128;  Gaps   20;

Qy         97 YRNI--DPIN-WFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNR 153
              | ||  : || ||:  :|::          | ::     ||  |::||
Db         43 YINIYLNYINLWFIYFMGLIVFF------LIFLLSRKLVSYSKYFYILL---------- 85

Qy        154 IHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDD-------FEVL 206
                              :|:|:   : ::|:                  ||       || |
Db         86 ----------------SYIFIPFDVTLIIL----------------IDDFMCFMILFESL 113

Qy        207 YYQISGHSLPLKLRQKAVYIAIVLPILSVLS---------VVITHVTMSDLNINQVVPYC 257
              ::  |   ||       : ::     | | ||          ::| |  :  ||:  :  |
Db        114 FFPICFVSLFFNFNNRFIFAIFYLVIFSSLSSSIMCIIICIIIIFHFNI--LNLQSFIDIC 171

Qy        258 ILDNLTAMLGAW-WFLICEAMSITAHLLAERFQKALKHIG-PAAMVADYRVLWLRLSKLT 315
              | |:|   || : | |:    ||        ::     |:  |   |     | : |: :
Db        172 IFDSL--YLGLYVWILLFIMFSI-------KYPIWPFHVWLPELHVEVNTELSILLASVV 222

Qy        316 RDTGNALCYTFVFMSL------YLFFIITLSIYGL------------MSQLSEGFGIKDI 357
                 |     | |:|:|       :| || :| : ||               ::   : :
Db        223 LKIGFFGLYKFLFLSFNQLSIWFLGFIDSLVMLGLTFLAITLLFLSDYKKIIATWSVIHT 282

Qy        358 GLTITALWN-----IGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFL 412
              |: :  ||:      :||| : |: :|  |           : : : :|   :     | : |
Db        283 GIGLILLWHNDIIFLGLLIF-CNLSHIIS--------SAFMFMMVGYMYDNYGVRIFLML 333

Qy        413 RATEMNPSTINCGGFFDVN--RTLFKGLLTTMVTYLVVLLQFQISI 456
               :            || ::   :|| |:    : :  :|| | | |
Db        334 IS------------FFGISIWSSLFLGIFLFNIDFPFMLL-FYIDI 366


RESULT 4
I64248
hypothetical protein homolog MG443 - Mycoplasma genitalium
C;Species: Mycoplasma genitalium
C;Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: I64248
R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M
, C.A.; Venter, J.C.
Science 270, 397-403, 1995
A;Title: The minimal gene complement of Mycoplasma genitalium.
A;Reference number: A64200; MUID:96026346; PMID:7569993
A;Accession: I64248
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-395 <TIGR>
A;Cross-references: UNIPROT:P47681; GB:U39731; GB:L43967; NID:g3845031; PIDN:AAC72463.1;
A;Experimental source: strain G-37
C;Genetics:
A;Genetic code: SGC3
C;Superfamily: Mycoplasma pneumoniae hypothetical protein yfiB

  Query Match             4.2%;  Score 111.5;  DB 2;  Length 395;
  Best Local Similarity   20.9%;  Pred. No. 0.22;
  Matches   78;  Conservative   63;  Mismatches  123;  Indels  109;  Gaps   19;

Qy        163 PFEEAVIAYLFLVN--ILPIMIIPILWYE----------ARKIAKLFNDWDD------FE 204
              | :  ::  | ::|  :| | |      |            || :  |   :|:      |
Db         41 PLKYVLVYLLSIINAFLLLIFIQKTGLYSFGISSLTQGFARLVFVLLKSFDETQRLLIFN 100
```

```
Qy        205 VLYYQISGHSLPLKLRQKAVYIAIVLPIL--------SVLSVVITHVTMSDLNINQVVPY 256
              :||:               :|: | :|::          ::: ||  ::    : |  :
Db        101 ILYW-------------LLYVPINIPLIIFSYKKIGKNFTILSTHFVVA----SNVFGF 142

Qy        257 CI-----LDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRL 311
               |      |||  ||                 | :    | || | :  :|    :  ||
Db        143 LISIIPGSDNLPPML-------------ASITDTNFWKAAKDLNQSAGFVPF--LWSDT 186

Qy        312 SKLTRDTGNALCYTFVFMSLYLFFI-ITLSIYGLMSQLSEGFGI----------KDIGLT 360
              |:     || :  ||:: ::| |:  |::|:  ::   : |                : :|
Db        187 SQ-----GNVIISTFIYAAIYGFYNGISVSLLYILGGSAGGADFLTQYYARKKNRSVG-- 239

Qy        361 ITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPS 420
                     :|||:       :: : :     ||: ::|    |:  |:::|      :|:
Db        240 -------SILFYVNSFILIIAILIGSFVAGSLLLQDVN-NYRDSAWEVSLF-----FSPN 286

Qy        421 TINCGGFFDVNRTLFKGLLT-TMVTYLVVLLQF-QISIPTDKGDSEGANNITVVDFVMDS 478
               |          | |  ||| |:|:||     | :| : |||   |       : |    |
Db        287 LI---------ATFFSILLTGTVVSYLFPRYNFAEIKVPTDK--LEEVRKALLSDNANHS 335

Qy        479 LDNDMSLMGASTL 491
              |    :| | | |
Db        336 LSIQETLGGYSLL 348
```

RESULT 5
H82509
hypothetical protein VCA0023 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: H82509
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: H82509
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-346 <HEI>
A;Cross-references: UNIPROT:Q9KNE2; GB:AE004346; GB:AE003853; NID:g9657401; PIDN:AAF9593
A;Experimental source: serogroup O1; strain N16961; biotype El Tor
C;Genetics:
A;Gene: VCA0023
A;Map position: 2

  Query Match 4.1%; Score 107.5; DB 2; Length 346;
  Best Local Similarity 21.2%; Pred. No. 0.4;
  Matches 75; Conservative 49; Mismatches 127; Indels 103; Gaps 16;

```
Qy        168 VIAYLFLVNILPIMIIPILWY--EARKIAKLFNDWDDFEV-LYYQISGHSLPLKLRQKAV 224
              ||| | ::  |   :    |: :   :| |||    | | |:: |||: :  ||    :
Db         20 VIAMLAII-ALHCQLFTTYWFLDDEPWVAYLFNQSTRFAVPLFFLISGYLIQPKLSHNPM 78

Qy        225 -----YIAIVLPILSVLSVVITHVTMSDLNINQVVPYCILDNLTAML---------GAWW 270
                  | : :| |  : ||           |: ::|:    ||  |:          | |
Db         79 QTLRNYCSPLLRIWVIWSV----------ISLLMPF----NLEVMVNQGYLAERSGYWG 123

Qy        271 FLICEAMSITAHLLAERFQKALKHIG--PAAMVA---------DYRVLW-----LRLSKL 314
              ||:        | |   |:  | |:   || |:|          :  |      : |
Db        124 FLL-------QHPLNSLFEGGLVHLWFLPALMIAVAIMALLIRQQKTHWMLPIAIGLYLY 176

Qy        315 TRDTGNALCYTFVFMSLY---------LFFII----------------TLSIYGLMSQL 348
                |::   | :   :|          || ::                  |:: |:
Db        177 GEFAGSSAVVTGMSAPIYTRNGPFFSTLFVVVGYLIRERHILWQSRSALLLAMLGMAFHF 236

Qy        349 SEGFGIKDIG--------LTITALWNIGLLFYICDEAH-------YASVNVRTNFQKKLL 393
                | :|:   |         |  | || |||  ::  :          ::       |    |
Db        237 VEAYGLHQYGQVFNTNDYLFGTTLWAIGLFLFLLAKPDLGRKPWGFSLSQSILGFYVSHL 296

Qy        394 MVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLV 447
```

```
                :| :  ||      :  ||    :   |:   |        | |||  | : :|:
Db        297 LVVIMMMN------LARFLGLAGLEKDTLVLFGTLLTTYLLVKGLERTPLKHLL 344
```

RESULT 6
AI1510
C-terminal domain glycerophosphoryl diester phosphodiesterase homolog lin0625 [imported]
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AI1510
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker .; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H. D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AI1510
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-583 <GLA>
A;Cross-references: UNIPROT:Q92E36; GB:AL592022; PIDN:CAC95857.1; PID:g16413065; GSPDB:GN
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0625

  Query Match 4.1%; Score 107.5; DB 2; Length 583;
  Best Local Similarity 21.4%; Pred. No. 0.76;
  Matches 84; Conservative 66; Mismatches 104; Indels 139; Gaps 24;

```
Qy        106 FLRIIGVLPIVRH---GPARAKPEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSG 162
              :|::  :| |::    ||          |||   ||:|    |  : :  :       |
Db         18 YLKVTLLLTILQAFVIGP----------FIYYFFFFILRVIGVPGITDANL-------G 59

Qy        163 PFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEVL--YYQISGHSLPL--- 217
                : :| : |: || ::|:  ::||             | ::   ||: | |  :
Db         60 EVFSSPVAVIMLL-ILALLILLFVYYEL----------GFFIMMAIYQLRGESYTVFKI 107

Qy        218 --KLRQKA-----------VYIAIVLPILSVLSVVITHVTMSDLNINQVVPYCILDNL- 262
                :|  ||           :|  ::|||   ||: ||       :  |  :|: | | |
Db        108 IQRLNVKAKYFLSYQAIYFLLYFFLLLPIAG-LSLPIT------ITENLYLPHFITDELM 160

Qy        263 TAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNAL 322
                | | ::|  |::|  :: | |   || :      : |     |::|         |:
Db        161 KTTTGTWLYVI--AIAIIFYISA-RLVFALPY-----FIEDKS---LKIS-------GAI 202

Qy        323 CYTFVFMSLYLFF-----IITLSIYG--------------LMSQLSEGFGIKDIGLTIT 362
              :: |   :|||     :: : : |                |: ::: |  :   |:|:|
Db        203 RKSWKFSQKHLFFMLLKWVLIIVVIGFLVSIIATIIMLPLLLVEKITPGIAVVIAGITLT 262

Qy        363 ALWNIGLL----------------FYICDEAHYASVNVRTNF--QKKLLMVELNWMNSD 403
                | ||                | |  |  :| |  |  |  :|: ::|
Db        263 ILQVIGFFAAGIFQGIIAQLLVKNAFAI--EGQHAPV-ARNQFPHKKRFIIV-------- 311

Qy        404 AQTEINMFLRATEMNPSTINCGGFFDVNRTLFK 436
                | :|:  :  |          | || ||::
Db        312 ---AIIVFIIFSGFN--------IFTVNATLYE 333
```

RESULT 7
D95228
hypothetical protein SP1953 [imported] - Streptococcus pneumoniae (strain TIGR4)
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C;Accession: D95228
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916

A;Accession: D95228
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-698 <KUR>
A;Cross-references: UNIPROT:Q97NR4; GB:AE005672; PIDN:AAK76021.1; PID:g14973459; GSPDB:G
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1953

Query Match 4.0%; Score 106.5; DB 2; Length 698;
Best Local Similarity 19.1%; Pred. No. 1.1;
Matches 128; Conservative 84; Mismatches 175; Indels 283; Gaps 30;

```
Qy          6 EKVVKGH------GQGNSGHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYS 59
              ||: |          :|    |:|    :|       : |  | |:|       :  ||
Db         88 EKISKNKFYILDPAKGRQRMSISEFERHYSN-----IILTFKKLDS-------FMSRK-- 133

Qy         60 IGLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLR-----IIGVLP 114
                          | |||         |  |    |||   | :|:      |  ::|
Db        134 ---------------DNKKS--------PVLKYFFKYRNKLGILFFVTALLYVIQSLVP 169

Qy        115 IVRHGPARAKFEMNSAS--------FIYSVVF---FVLLACYVG---------------- 147
              |          |: :| |         ||::| |   ::|   ||
Db        170 IANRYIIDTNFKDDSYSSRMLFTILFIFTVSFSLMYLLRQIYVASLKYIMDKEISYDFMK 229

Qy        148 -------------------YVANNRIHIVRSLSGPFEEAVIAYL---------------- 172
                              : ||: ::|   ||  |  |::  |
Db        230 HLIYLPYSFYEKRTLGDILFRANSIVYIREILSNNFIAAILDLLMIVVYAVVLFSFSKYM 289

Qy        173 --FLVNI---LPIMIIPILWYEARKIAK-----------------------------LF 197
                ||:::   | |:: ||:      | |
Db        290 VIFLISLSLALSIVMYPIIKISKNLIDKNIKEKVNVQNITSEVISKNSDIKLTGEEEFWI 349

Qy        198 NDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTM---------SDL 248
              | ||:|      |          |:  ::::||  | :|| :::  :|:           |
Db        350 NKWDNFNTKQLIIG--------RKLDIHLSIVSSITNVLQIILPVLTLIVGVNIKTFEQL 401

Qy        249 NINQVV------PYCI------LDN-LTAMLGAWWFL-ICEAMSITAHLLAERFQKALKH 294
              : |:|      || |       || :  ||   :|| | :  :  : |: ||  : :|
Db        402 TLGQIVAISTVSPYFISPIISLSDNYIQLMLLKGYFLRIEDVFNTKSELIPERVSQDIKF 461

Qy        295 IGPAAMVADYRV----LWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSE 350
                        | ::   :| :                                 |||   :
Db        462 --------DKKIELKDIWYK-----------------------------YGLF----D 478

Qy        351 GFGIKDIGLTI-----------------TALWNIGLLFYICDEAHYASVNV-------- 384
              : :| | :||                  |   :|||     | :  |: |
Db        479 DYVLKGINVTIKKGETVAIVGESGSGKSTLAKILLGLL-----EPNIGSIEVDGVEKEEI 533

Qy        385 -RTNFQKKLLMVELNWMNSDAQTEINM----FLRATEMNPSTINCGGFFDVNRTLFKGLL 439
               :| ::|    |  |   |       |:    |:   |:  : :|  |  :| ::|  ||
Db        534 GQTLYRKIFGAVLQNSTLSYGTLRENLTFGHFVSDEEL-MTNLNSIGLSNVVKSLPLGLE 592

Qy        440 TTMV----------TYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLDNDMSLMGAS 489
              | :              :::| :  :| |           :: |:|    ||||    :  |
Db        593 TIIAEEGNNFSGGQQQMILLARCLLSKP----------SVVVLDEATSSLDNLSQQITTS 642

Qy        490 TLSTTTVGTT 499
                ||   :|||
Db        643 YLS--EIGTT 650
```

RESULT 8
H98092
hypothetical protein clyB [imported] - Streptococcus pneumoniae (strain R6)
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: H98092
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: H98092
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-702 <KUR>
A;Cross-references: UNIPROT:Q8DNF4; GB:AE007317; PIDN:AAL00573.1; PID:g15459452; GSPDB:GN
C;Genetics:
A;Gene: clyB

Query Match 4.0%; Score 106.5; DB 2; Length 702;
Best Local Similarity 19.1%; Pred. No. 1.1;
Matches 128; Conservative 84; Mismatches 175; Indels 283; Gaps 30;

```
Qy          6 EKVVKGH------GQGNSGHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYS 59
              ||: |          :|    |:|    :|       : |  | |:|       :  ||
Db         92 EKISKNKFYILDPAKGRQRMSISEFERHYSN-----IILTFKKLDS-------FMSRK-- 137

Qy         60 IGLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLR-----IIGVLP 114
                          | |||         |  |    |||   | :|:      |  ::|
Db        138 ---------------DNKKS--------PVLKYFFKYRNKLGILFFVTALLYVIQSLVP 173

Qy        115 IVRHGPARAKFEMNSAS--------FIYSVVF---FVLLACYVG---------------- 147
              |          |: :| |         ||::| |   ::|   ||
Db        174 IANRYIIDTNFKDDSYSSRMLFTILFIFTVSFSLMYLLRQIYVASLKYIMDKEISYDFMK 233

Qy        148 -------------------YVANNRIHIVRSLSGPFEEAVIAYL---------------- 172
                              : ||: ::|   ||  |  |::  |
Db        234 HLIYLPYSFYEKRTLGDILFRANSIVYIREILSNNFIAAILDLLMIVVYAVVLFSFSKYM 293

Qy        173 --FLVNI---LPIMIIPILWYEARKIAK-----------------------------LF 197
                ||:::   | |:: ||:      | |
Db        294 VIFLISLSLALSIVMYPIIKISKNLIDKNIKEKVNVQNITSEVISKNSDIKLTGEEEFWI 353

Qy        198 NDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTM---------SDL 248
              | ||:|      |          |:  ::::||  | :|| :::  :|:           |
Db        354 NKWDNFNTKQLIIG--------RKLDIHLSIVSSITNVLQIILPVLTLIVGVNIKTFEQL 405

Qy        249 NINQVV------PYCI------LDN-LTAMLGAWWFL-ICEAMSITAHLLAERFQKALKH 294
              : |:|      || |       || :  ||   :|| | :  :  : |: ||  : :|
Db        406 TLGQIVAISTVSPYFISPIISLSDNYIQLMLLKGYFLRIEDVFNTKSELIPERVSQDIKF 465

Qy        295 IGPAAMVADYRV----LWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSE 350
                        | ::   :| :                                 |||   :
Db        466 --------DKKIELKDIWYK-----------------------------YGLF----D 482

Qy        351 GFGIKDIGLTI-----------------TALWNIGLLFYICDEAHYASVNV-------- 384
              : :| | :||                  |   :|||     | :  |: |
Db        483 DYVLKGINVTIKKGETVAIVGESGSGKSTLAKILLGLL-----EPNIGSIEVDGVEKEEI 537

Qy        385 -RTNFQKKLLMVELNWMNSDAQTEINM----FLRATEMNPSTINCGGFFDVNRTLFKGLL 439
               :| ::|    |  |   |       |:    |:   |:  : :|  |  :| ::|  ||
Db        538 GQTLYRKIFGAVLQNSTLSYGTLRENLTFGHFVSDEEL-MTNLNSIGLSNVVKSLPLGLE 596

Qy        440 TTMV----------TYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLDNDMSLMGAS 489
              | :              :::| :  :| |           :: |:|    ||||    :  |
Db        597 TIIAEEGNNFSGGQQQMILLARCLLSKP----------SVVVLDEATSSLDNLSQQITTS 646

Qy        490 TLSTTTVGTT 499
                ||   :|||
Db        647 YLS--EIGTT 654
```

RESULT 9
B71640
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain L1 RP792 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 09-Jul-2004
C;Accession: B71640
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U.
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499; PMID:9823893

```
A;Accession: B71640
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-653 <AND>
A;Cross-references: UNIPROT:Q9ZCG1; GB:AJ235273; GB:AJ235269; NID:g3861237; PIDN:CAA1521
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: nuoL1; RP792
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: membrane-associated complex; NAD; oxidoreductase

  Query Match             4.0%;  Score 105.5;  DB 2;  Length 653;
  Best Local Similarity   20.3%;  Pred. No. 1.3;
  Matches   95;  Conservative   76;  Mismatches  178;  Indels  119;  Gaps   24;

Qy         93 PSVFYRNIDPINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVPFVLLACYVGYVANN 152
Db         68 PWIEFKNLQ-VNWSIYIDQLTSIM----------FIAVTFVSSVVHIYSL----GYMAED 112

Qy        153 RIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEVLYYQISG 212
Db        113 K-GIIRFLS---------FLSLFTFFMLMLV-----SADNFLQLFCGWEGVGVCSYLLIG 157

Qy        213 HSLPLKLRQKAVYIAIVLP-------ILSVLSVVITH--VTMSDL--------NINQVVP 255
Db        158 FWHSKESANKAAIKAFITNRVSDFAFILGIITIIVYYGSANYKDVFSSAKLLSNTKIFVH 217

Qy        256 YCILDNLTAMLGAWWFLIC--EAMSITAHL-LAERFQ-----KALKHIGPAAMVADYRVL 307
Db        218 FSILDIICLLL----FIGCMGKSAQIGLHVWLPDAMEGPTPVSALIH--AATMVTAGVFL 271

Qy        308 WLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGI--KDIG--LTITA 363
Db        272 VARCS-----------YLFEYSPIVLQFITIIG--GITCLFAASIAIMQSDIKKIIAYST 318

Qy        364 LWNIGLLFYICDEAHYASV---NVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPS 420
Db        319 CSQLGYMFMACGVSSYNSAIFHLVTHAFFKALLFLSAGNV-IHAVNEHNIFKMGGLINKM 377

Qy        421 TINCG----------------GFFDVNRTL-----------FKGLLTTMVTYL----VVL 449
Db        378 PITYGNFLIGSLALIGIYPLSGFYSKDLILEATYSSGSFMFIFGIITAILTAIYSMKIII 437

Qy        450 LQFQISIPTDKGDSEGANNITVVDFVMDSLDNDMSLMGASTLSTTTVG 497
Db        438 LVFHGKTKLEKDVFKHAHEPTKI------MNNPLILLVAGSFFSGMIG 479


RESULT 10
H71732
cytochrome D ubiquinol oxidase chain I (cydA) RP216 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii
C;Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 09-Jul-2004
C;Accession: H71732
R;Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.; Alsmark, U
Nature 396, 133-140, 1998
A;Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A;Reference number: A71630; MUID:99039499; PMID:9823893
A;Accession: H71732
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-453 <AND>
A;Cross-references: UNIPROT:Q9ZDV3; GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA1467
A;Experimental source: strain Madrid E
C;Genetics:
A;Gene: cydA; RP216
C;Superfamily: cytochrome d complex terminal oxidase chain I

  Query Match             4.0%;  Score 105;  DB 2;  Length 453;
  Best Local Similarity   23.4%;  Pred. No. 0.89;
  Matches   68;  Conservative   50;  Mismatches   98;  Indels   74;  Gaps   17;

Qy         96 FYRNIDPINWFLRIIGVLPIVRHG--PARAKFEMNSASFIYSVVPFVLLACYVGYVANNR 153
Db        166 FY----PLNW-------LEIIFNPSFPYRF-FHMITAS--YLTTSFVIGGVASFYLLNNR 211

Qy        154 IHIVRSLSGPFEEAVIAYLFLVNILPIMIIPI-----------LWYEARKIAKLFNDWD 201
Db        212 ----------YKKHAKIMLFMAVLMALIVSPIQIFIGDLHGLNTLKYQPVKVSAIEGIWN 261

Qy        202 DFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVT------MSDLNINQVVP 255
Db        262 TEKGASFNLIG--LPDKEEEKTKY-AIEIPYAS--SLILTHSLDGEVKGLKEWTKEERPP 316

Qy        256 YCIL-DNLTAMLGAWWFLICEAMSITAHLLAER------FQKALKHIGPAAMVADYRVL- 307
Db        317 VAVVFFSFRIMLGIGCLMVFTGIAGLYLYLNKRLYTTYWFQYWYILMSPSGPIA---VLA 373

Qy        308 -WL-----------RLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGL 344
Db        374 GWLVTEVGRQPYIVYNILK-TVDTVSPLLGKYVFISLIAFVVVYLIIFGV 422


RESULT 11
A75145
protein translocase chain (secy) PAB2139 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: A75145
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome struc
A;Reference number: A75001
A;Accession: A75145
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-468 <KAW>
A;Cross-references: UNIPROT:Q9V1V8; GB:AJ248284; GB:AL096836; NID:g5457730; PIDN:CAB4924(
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB2139
C;Superfamily: yeast SSH1 protein

  Query Match             4.0%;  Score 105;  DB 2;  Length 468;
  Best Local Similarity   21.9%;  Pred. No. 0.92;
  Matches   68;  Conservative   53;  Mismatches  121;  Indels   68;  Gaps   13;

Qy         84 TASNNPEFKPSVFYRNIDPINWFLRIIGVLP-----IVRHGPARAKFEMNSASFIYSVVF 138
Db        191 TRSLNPLTDPNI----IDPLTGKPAIVGAIPYFIQHILKGDLKGALYRGGTAPDMMAVIA 246

Qy        139 FVLLACYVGYVANNRIHIVRSLSG-------PFEEAVIAYLFLVNILPIMIIPI-----L 186
Db        247 TIIVFLVVVYFESMRVEIPLGYRGVTIRGRYP-----IRFLYVSNIPIILTFALYANIQL 301

Qy        187 WYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTMS 246
Db        302 W--ARVLDRLGHPW----------LGTFDPTTGNPVGGFVLYVIPPRSIFTVIDNPV--- 346

Qy        247 DLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRV 306
Db        347 -----RALVYLILTVISSLI--FGFLWVELTGLDARTIARQLQRAGLQI--PGFRRDPRT 397

Qy        307 LWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWN 366
Db        398 LERVLQKY-------IPYVTFWGSLTVALIAVLADF--LGALGTGTGI---------LLT 439

Qy        367 IGLLFYICDE 376
Db        440 VGILYRFYEE 449


RESULT 12
E64245
periplasmic phosphate permease homolog AG88 homolog - Mycoplasma genitalium
C;Species: Mycoplasma genitalium
```

```
C;Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 09-Jul-2004
C;Accession: E64245
R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.
, C.A.; Venter, J.C.
Science 270, 397-403, 1995
A;Title: The minimal gene complement of Mycoplasma genitalium.
A;Reference number: A64200; MUID:96026346; PMID:7569993
A;Accession: E64245
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-654 <TIGR>
A;Cross-references: UNIPROT:P47651; GB:U39726; GB:L43967; NID:g1046113; PID:g1046125; TI
A;Experimental source: strain G-37
C;Genetics:
A;Genetic code: SGC3
A;Start codon: GTG
C;Superfamily: periplasmic phosphate permease AG88

  Query Match             4.0%;  Score 105;  DB 2;  Length 654;
  Best Local Similarity   21.3%;  Pred. No. 1.4;
  Matches  104;  Conservative   73;  Mismatches  153;  Indels  158;  Gaps   27;

Qy         89 PEFKPSVFYRNIDPIN-----WFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLA 143
              |:|  |:|    :  |     ||       |::             |||  | :  :::|
Db         46 PDFAKSLFNLEFNLGNKQAGIWF-------PLL-------------VSFIVS-IGALIIA 84

Qy        144 CYVG--------YVANNRIH-----IVRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEA 190
              |:|          |    :|      |:  |||    :||  ||   || |
Db         85 SYIGVRTSFFLVYRCKPKIRKKLSLIIDILSG--IPSVIFGLFASQILSIFF-------- 134

Qy        191 RKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIA-IVLPIL-------------SVL 236
              | | ||                     || |      :: :::||:             ::
Db        135 RDILKL------------------PPLSLLNVIAMLSFMIIPIVISLTTNTLTYVNNDLI 176

Qy        237 SVVITHVTMSDLNINQVVPYCILDNLTAMLG-AWWFLICEAMSITAHLLAERFQKALKHI 295
              |||::         | :::    |   || :|  |:   | | |::   | :  :|: : :
Db        177 SVVVSLGENKTSAIYKIIKKEIKPQLTVILTLAFARAISETMAVNFVLQSVNYQEVINN- 235

Qy        296 GPAAMVADYRVLWLRLSK-LTRDTG----NALCYTF-----VFMSLYLFFII------TL 339
                 :| : |   :|  :  : |    | : | |      : :||  || |       ||
Db        236 -NRFFTSDLKTLGSVISTFIFSENGDEQINGVLYIFGIIILILVSLLNFFAIWSANPKTL 294

Qy        340 SIYGLMSQLSEGFGIKDIGLTITALW----NIGLLFYICDEAHYA-------SVNVRTNF 388
                | : ::|             :|    ||  ||        :       ::| |: |
Db        295 ERYPFLKKIS--------NFIYQVVWFIPNNISALFVDLTSTRQSVKKIKVNNINERSLF 346

Qy        389 QKKLLM----VELNWMNSDAQTEINMFLR-----------------ATEMNPSTINCGGF 427
               |: |     ::||:     |  |  ||                   |  | ||:
Db        347 FKERLQSVVWIKLNYFLKIFQELICTFLAFGFVLAILLFVFINGSVAINNNGSTVFS--- 403

Qy        428 FDVNRTLFKGLLTTMVTYLVVLLQFQISIP---------TDKGDSEGANNITVVDFVMDS 478
              |: : |  : |: |:|   ::|:   |: |         :  :|:   |  | :||:||
Db        404 FEADST-GRALVNTLV---IILITITITFPLALLIAIWLNEYNNSKVVKN--VFNFVIDS 457

Qy        479 LDNDMSLM 486
              | :  |::
Db        458 LSSMPSII 465


RESULT 13
B75059
probable transmembrane oligosaccharyl transferase PAB0974 - Pyrococcus abyssi (strain Or
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C;Accession: B75059
R;anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Reference number: A75001
A;Accession: B75059
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-766 <KAW>
A;Cross-references: UNIPROT:Q9UYP5; GB:AJ248287; GB:AL096836; NID:g5458657; PIDN:CAB50367
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: PAB0974

  Query Match             4.0%;  Score 105;  DB 2;  Length 766;
  Best Local Similarity   20.1%;  Pred. No. 1.7;
  Matches  111;  Conservative   70;  Mismatches  188;  Indels  182;  Gaps   24;

Qy         18 GHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSIGLAERLD-----ADYEA 72
              ||    ||||||   ||             |  : | :    |  | |         |
Db        139 GHIFRSMANYYR---GD-----------NYMLFWYSLALAGIAYALRTKGYRRLVFYLV 183

Qy         73 PPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGV-LPIVRHGPARAKFEMNSAS 131
              | |    |              |||::   |: || : || | |      : :  ::|
Db        184 PTLASGIS-------------SVFWQAYYPLFVFLSLNGVFLAIGSFLLDKKRNFLDSFI 230

Qy        132 FIYSVVFFVLLACYV------GYVANNRIHIVRSLSGPFEEAVIAYLF--LVNILPIM-- 181
              | |  |  ::| |:      | :  || |||  |     :   ||||  |  ::||
Db        231 IILSTAFGAIIANYLGEKFGYGMLGYNRQHIVSKLGIKLGKIRDAYLFIHLHYLVPISLG 290

Qy        182 IIPILWYEAR----KIAKL---------------------------------------- 196
              :: || : :|    | ||:
Db        291 LLIILLFLSRFVKSKRAKVGIVIGLGTVSILIILLKFPALRGLLGIFDMFKSTPIMETRP 350

Qy        197 --FND-WDDFEV-------------------------------LYY-------------- 208
                |:| |  | :                              ||
Db        351 TNFHDLWKAFSISIFLLPLFFLRFHPEKVKTEDFFLLGLIVPSLYMLLAWARFVFVGSLA 410

Qy        209 --------QISGHSLPLKLRQ--KAVYIAIVLPILSVLSVVITHVTMSDL-----NINQV 253
                     : |:|| :: |:  ||  :|::| ||  | :|    |: :        ||:
Db        411 VATMAGIGLVEGYSLVIQRRKSGKASRVALILLIL--LLIVNGAFTLKNCLSMRPLINKE 468

Qy        254 VPYCIL-----DNLTAMLGAWWFL---------------ICEAMSITAHLLAERFQKALK 293
                 ::       |   :: |||                     |      :  : |  | :  :
Db        469 WENALIWLKNNSNENDVILAWWDYGAWITYYSRRAPVAEIAPNPDVALYYLGARNRDWIM 528

Qy        294 HIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFG 353
               :|   ::  |   :|: | : |        |  :          :| |  |::  : ||
Db        529 SLGVDYVIVSY-YDFLKFSSIVRTASAQSGYNLIKRYWIAVLPLTSSYGGIL--IFEGGE 585

Qy        354 IKDIGLTITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINM-FL 412
               | |      :|:: ::  | |   |    :   :: |:   :| : ::||    ||: :
Db        586 YKIIAKP-GDIWDVRII--IGDHVVYPR-GLYVEYKGKVTESKLKYSSTDAYLYINLNYK 641

Qy        413 RATEMNPSTIN 423
              |  ||  | |
Db        642 YAILMNSETFN 652


RESULT 14
H86762
ABC transporter permease protein ylbB [imported] - Lactococcus lactis subsp. lactis (stra
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C;Accession: H86762
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlic
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: H86762
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-896 <STO>
A;Cross-references: UNIPROT:Q9CGJ2; GB:AE005176; PID:g12724061; PIDN:AAK05202.1; GSPDB:GN
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: ylbB

  Query Match             4.0%;  Score 104.5;  DB 2;  Length 896;
  Best Local Similarity   22.3%;  Pred. No. 2.3;
  Matches   63;  Conservative   42;  Mismatches  111;  Indels   67;  Gaps   13;
```

```
Qy        157 VRSLSGPFEEAVIAYLFLVNILPIM-----------IIPILWYEARKIAKLFNDW----D 201
Db        361 IDSLSNIFPVVLFAIALLVSLTTMTRFVEEERGNLGLLKALGYSNRDIRKKFMVYGLVSS 420

Qy        202 DFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTMSDLNIN-----QVVPY 256
Db        421 GLGALVGTIIGHTF--------------LPI-AVFNAYTASSTFSNLRLTLSPLWTIVAF 465

Qy        257 CILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTR 316
Db        466 AIA-IACSLLPAYWVVRMELKEVPASLFLAKVPKA----GSRILLEKINFIWKRMSFTYK 520

Qy        317 DTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWN--IGLLFYIC 374
Db        521 VTARNL---FRYKKRMLMTI--FGVAGCTALLVMGFGIRD---SISGLSNKQFGQIL--- 569

Qy        375 DEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEM 417
Db        570 ---HY-----------DMITIEKNKVNDKEKEEIDKELASSEI 598


RESULT 15
G70007
conserved hypothetical protein yueF - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: G70007
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: G70007
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-369 <KUN>
A;Cross-references: UNIPROT:O32095; GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15168.
A;Experimental source: strain 168
C;Genetics:
A;Gene: yueF
C;Superfamily: Bacillus subtilis conserved hypothetical protein yueF

  Query Match             3.9%;  Score 104;  DB 1;  Length 369;
  Best Local Similarity   18.7%;  Pred. No. 0.83;
  Matches   83;  Conservative   65;  Mismatches  139;  Indels  156;  Gaps   19;

Qy        105 WFLRIIGVLPIVRHGPARAKFEMNSASFIY-------SVVFFVLLACYVGYVANNRI--- 154
Db          9 WTLQILFVLLII--------FVATKVSFVFQPFIVFISTLFFPMLIAGILYFIFNPVVRL 60

Qy        155 ---HIVRSLSGPFEEAVIAYLFLVNILPIM---IIPILWYEARKIAKLFNDWDDF----E 204
Db         61 LEKKIPRTLS-----ILLIYLLFIGLLAFISASVGPII---TAQVTGLFNNLPDYIKQIQ 112

Qy        205 VLYYQISGHSLPLKLRQKAVYIAI----------------VLPILSVLSVVITHVTMSD 247
Db        113 ALTKDLS-HSQWFTWMMNQDYVSISKIEQSLTSFLQNLPQNITSSLSAVFGVVTNITLVI 171

Qy        248 LNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVL 307
Db        172 ITVPFILFYMLKD----------------------GHRF----PHLAVKILPASYRTE 203

Qy        308 WLRLSKLTRDT------GNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTI 361
Db        204 GLKIFKDLSDTLAAYFQGQLLICLFVGTACFIGYLIAGLPYALI---------------- 247

Qy        362 TALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPST 421
Db        248 -----LGIVMAITNIIPY----------------------------VGPFLGAA---PAV 271

Qy        422 INCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLDN 481
Db        272 I--VGFMD---SPAKALFAIIVVVIVQQLDGNLLSPLVIGKRLNTHPLTII--------- 317

Qy        482 DMSLMGASTLSTTTVGTTLPPPI 504
Db        318 -LLLIGAGSFG-GILGMILAVPV 338


Search completed: May 23, 2005, 19:25:56
Job time : 44 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 23, 2005, 18:58:23 ; Search time 171 Seconds
(without alignments)
1533.243 Million cell updates/sec

Title: US-10-081-816-12
Perfect score: 2645
Sequence: 1 MRPSGEKVVKGHGQGNSGHS.....TTTVGTTLPPPIMKLKGRKG 512

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2645	100.0	512	1	G63A_DROME	Q9vzl7 drosophila
2	1276	48.2	379	2	Q7PYF4	Q7pyf4 anopheles g
3	480	18.1	485	2	Q7PQT4	Q7pqt4 anopheles g
4	458	17.3	467	2	Q7PMG3	Q7pmg3 anopheles g
5	405	15.3	447	1	G21A_DROME	Q9vpt1 drosophila
6	167.5	6.3	412	1	G98D_DROME	Q8imn5 drosophila
7	149.5	5.7	383	2	Q7PVK1	Q7pvk1 anopheles g
8	148.5	5.6	472	2	Q7PEV2	Q7pev2 anopheles g
9	145.5	5.5	404	2	Q7PVJ9	Q7pvj9 anopheles g
10	143	5.4	444	2	Q7PPB5	Q7ppb5 anopheles g
11	142	5.4	416	2	Q7PKI4	Q7pki4 anopheles g
12	141.5	5.3	411	2	Q7PVK0	Q7pvk0 anopheles g
13	140	5.3	382	2	Q7PVJ8	Q7pvj8 anopheles g
14	139.5	5.3	451	2	Q7PKC5	Q7pkc5 anopheles g
15	139.5	5.3	1210	2	Q7QKL0	Q7qkl0 anopheles g
16	135.5	5.1	369	1	G39B_DROME	P58960 drosophila
17	132.5	5.0	406	1	G64B_DROME	P83294 drosophila
18	130.5	4.9	372	1	G391_DROME	P58956 drosophila
19	130.5	4.9	393	2	Q7PK41	Q7pk41 anopheles g
20	130	4.9	403	2	Q7PK35	Q7pk35 anopheles g
21	129	4.9	436	1	G61A_DROME	Q9w0m2 drosophila
22	127	4.8	388	2	Q7PK39	Q7pk39 anopheles g
23	126	4.8	390	2	Q8REF8	Q8ref8 fusobacteri
24	126	4.8	396	2	Q7PK37	Q7pk37 anopheles g
25	125.5	4.7	386	1	G22B_DROME	P84180 drosophila
26	125	4.7	399	2	Q7PK31	Q7pk31 anopheles g
27	125	4.7	403	1	G98B_DROME	Q9vb26 drosophila
28	125	4.7	451	1	G64E_DROME	P83296 drosophila
29	125	4.7	460	2	Q7KV53	Q7kv53 drosophila
30	124	4.7	401	2	Q8MMH7	Q8mmh7 heliothis v
31	124	4.7	419	2	Q9MD52	Q9md52 phytomonas
32	123.5	4.7	400	2	Q7PK34	Q7pk34 anopheles g
33	123	4.7	405	2	Q7PK33	Q7pk33 anopheles g
34	122.5	4.6	404	2	Q7PKI5	Q7pki5 anopheles g
35	122	4.6	414	1	G02A_DROME	Q9w594 drosophila
36	121.5	4.6	456	1	G64A_DROME	P83293 drosophila
37	121	4.6	397	2	Q7PFJ8	Q7pfj8 anopheles g
38	120.5	4.6	399	2	Q7PK38	Q7pk38 anopheles g
39	120	4.5	384	2	Q833X6	Q833x6 enterococcu
40	118.5	4.5	4099	2	Q7RPL9	Q7rpl9 plasmodium
41	118	4.5	396	2	Q7PK36	Q7pk36 anopheles g
42	117.5	4.4	445	2	Q7PF20	Q7pf20 anopheles g
43	117	4.4	429	2	Q7PI94	Q7pi94 anopheles g
44	117	4.4	572	2	Q6JV99	Q6jv99 bombyx mand
45	116	4.4	397	2	Q7PVJ7	Q7pvj7 anopheles g

ALIGNMENTS

RESULT 1
G63A_DROME
ID G63A_DROME STANDARD; PRT; 512 AA.
AC Q9VZL7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-JAN-2005 (Rel. 46, Last annotation update)
DE Putative gustatory receptor 63a.
GN Name=Gr63a; ORFNames=CG14979;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

```
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   GENOME REANNOTATION.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [3]
RP   IDENTIFICATION.
RX   MEDLINE=21407712; PubMed=11516643; DOI=10.1016/S0960-9822(01)00258-5;
RA   Dunipace L., Meister S., McNealy C., Amrein H.;
RT   "Spatially restricted expression of candidate taste receptors in the
RT   Drosophila gustatory system.";
RL   Curr. Biol. 11:822-835(2001).
CC   -!- FUNCTION: Probable role in the gustatory response.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Belongs to the G-protein coupled receptor Dr-tr
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE003479; AAF47803.1; -.
DR   FlyBase; FBgn0035468; Gr63a.
DR   GO; GO:0016021; C:integral to membrane; NAS.
DR   GO; GO:0008527; F:taste receptor activity; NAS.
DR   GO; GO:0050912; P:perception of taste, sensory transduction o. . .; NAS.
DR   GO; GO:0007607; P:taste perception; NAS.
KW   G-protein coupled receptor; Glycoprotein; Hypothetical protein;
KW   Multigene family; Receptor; Transmembrane.
FT   DOMAIN       1    129       Extracellular (Potential).
FT   TRANSMEM    130    150       1 (Potential).
FT   DOMAIN      151    166       Cytoplasmic (Potential).
FT   TRANSMEM    167    187       2 (Potential).
FT   DOMAIN      188    222       Extracellular (Potential).
FT   TRANSMEM    223    243       3 (Potential).
FT   DOMAIN      244    252       Cytoplasmic (Potential).
FT   TRANSMEM    253    273       4 (Potential).
FT   DOMAIN      274    324       Extracellular (Potential).
FT   TRANSMEM    325    345       5 (Potential).
FT   DOMAIN      346    350       Cytoplasmic (Potential).
FT   TRANSMEM    351    371       6 (Potential).
FT   DOMAIN      372    436       Extracellular (Potential).
FT   TRANSMEM    437    457       7 (Potential).
FT   DOMAIN      458    512       Cytoplasmic (Potential).
FT   CARBOHYD    431    431       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   512 AA;  57465 MW;  B98BD94D6D4BFFF6 CRC64;

  Query Match             100.0%;  Score 2645;  DB 1;  Length 512;
  Best Local Similarity   100.0%;  Pred. No. 5e-189;
  Matches  512;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRPSGEKVVKGHGQGNSGHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRPSGEKVVKGHGQGNSGHSLSGMANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSI 60

Qy         61 GLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGP 120

Qy        121 ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPI 180

Qy        181 MIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVI 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 MIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVI 240

Qy        241 THVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAM 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 THVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAM 300

Qy        301 VADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 VADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLT 360

Qy        361 ITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPS 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPS 420

Qy        421 TINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDSEGANNITVVDFVMDSLD 480

Qy        481 NDMSLMGASTLSTTTVGTTLPPPIMKLKGRKG 512
              ||||||||||||||||||||||||||||||||
Db        481 NDMSLMGASTLSTTTVGTTLPPPIMKLKGRKG 512


RESULT 2
Q7PYF4
ID   Q7PYF4      PRELIMINARY;      PRT;   379 AA.
AC   Q7PYF4;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   AgCP12374 (Fragment).
GN   Name=agCG53552; ORFNames=ENSANGG00000017642;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008987; EAA01019.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   379 AA;  42958 MW;  135438038754E0F2 CRC64;

  Query Match             48.2%;  Score 1276;  DB 2;  Length 379;
  Best Local Similarity   66.4%;  Pred. No. 5.3e-87;
  Matches  241;  Conservative   50;  Mismatches   72;  Indels    0;  Gaps    0;

Qy         97 YRNIDPINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHI 156
              : :  ||   || |||||  |       | : | |  | |:||:||  |: ::  ||| |
Db          4 FESSKPIYLVLRAIGVLPYTRLPSGGTAFVLASPSMTYCVLFFLLLTVYIAFILLNRIEI 63

Qy        157 VRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLP 216
              ||:| | |||:||||||:||||||:|||::|||::||:  : | | ||| :| : :| :|
Db         64 VRTLEGRFEESVIAYLFIVNILPILIIPLMWYESRKVVSVVNGWVDFETVYRETTGRALE 123

Qy        217 LKLRQKAVYIAIVLPILSVLSVVITHVTMSDLNINQVVPYCILDNLTAMLGAWWFLICEA 276
              |:|| ||  |||:|||| ||| |||||| |  : ||:|||:|| :| |:| :|:: ||
Db        124 LRLRTKAQVIAILLPILCSLSVAITHVTMVDFKLLQVIPYCVLDTITYMMGGYWYMACET 183

Qy        277 MSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFI 336
              :|||| :||| ||:||:|:|||| |::|| |||||||| |||| : |||| |: ||||||
Db        184 LSITAKILAEDFQRALRHVGPAAKVSEYRSLWLRLSKLARDTGFSTCYTFTFICLYLFFI 243
```

```
Qy        337 ITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVE 396
              ||||||||||:|:|||:||||| :||  ::|||||||||||||| ||||||||||||||
Db        244 ITLSIYGLMSQISDGFGVKDIGLAVTAFCSVGLLFYICDEAHYASFNVRTNFQKKLLMVE 303

Qy        397 LNWMNSDAQTEINMFLRATEMNPSTINCGGPFFDVNRTLFKGLLTTMVTYLVVLLQFQISI 456
              |:|||:|||||||||||||||||:|| |||||||||||| || ||||||||||||||||
Db        304 LSWMNTDAQTEINMFLRATEMNPSSINLGGPFFDVNRTLFKSLLATMVTYLVVLLQFQISI 363

Qy        457 PTD 459
              | :
Db        364 PDE 366


RESULT 3
Q7PQT4
ID   Q7PQT4      PRELIMINARY;      PRT;   485 AA.
AC   Q7PQT4;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000003178.
GN   Name=ENSANGG00000002573;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008879; EAA08342.2; -.
SQ   SEQUENCE   485 AA;  55676 MW;  8F30114775BFF5BF CRC64;

  Query Match             18.1%;  Score 480;  DB 2;  Length 485;
  Best Local Similarity   29.6%;  Pred. No. 1.8e-27;
  Matches  113;  Conservative   93;  Mismatches  170;  Indels    6;  Gaps    4;

Qy         96 FYRNIDPINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIH 155
              |||:   :    | : |:|| |  | |  |   ||: ||:  |:::    |  |   ||
Db        102 FYRDHKLLLVLFRGLAVMPITRSVPGRITFSWRSAASIYAFCFYLVSTVIVLVVGYERIK 161

Qy        156 IVRSLSGPFEEAVIAYLFLVNILPIMIIPIL-WYEARKIAKLFNDWDDFEVLYYQISGHS 214
              : :: :  |:| :   ||:: ::|   || : |  |:::|     |  |:| ||:::| |
Db        162 VFQTTT-KFDEYIYGILFVIFLVPHFWIPFVGWGVAKQVAIYKTMWGAFQVRYYRVTGTS 220

Qy        215 LPL-KLRQKAVYIAIVLPILSVLSVVITHVTMSDLNINQVVPYCILDNLTAMLGAWWFLI 273
              |    |:   |:::|   : ::: ::     :   :      |  :  :  |  | |::
Db        221 LQFPHLKLLIVFLSIGCLVCAIVFLLSLSFLLEGFALWHTSAYYHIITMLNMNSALWYIN 280

Qy        274 CEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSLYL 333
                : :  |:  |:: :     |||:: || ||| ||:| :  |||   |:     |::
Db        281 SRGIRVASSSLSRCFRQDVAIECTAAMISRYRFLWLNLSELLQALGNAYARTYSTYCLFM 340

Qy        334 FFIITLSIYGLMSQLSE---GFGIKDIGLTITALWNIGLLFYICDEAHYASVNVRTNFQK 390
              |  ||::||| :|:: :   ||  |:||| :  ::   |||  || :| |:: |    |
Db        341 FVNITVAIYGALSEIIDHGFGFSFKEIGLIVDTVYCSTLLFIFCDCSHNATLQVAQGVQD 400

Qy        391 KLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGPFFDVNRTLFKGLLTTMVTYLVVLL 450
              || : |  ::   | ||::|::| ||||: ::  |: :|||  |    : |:  ||:|||
Db        401 TLLSINLLKVDQPTQKEIDLFIQAIEMNPAIVSLKGYAEVNRELLTSSIATIAIYLIVLL 460

Qy        451 QFQISIPTDKGDSEGANNITVV 472
              ||::|: : :   |   |: ::
Db        461 QFKLSLISQQIPVEIIENVKLL 482


RESULT 4
Q7PMG3
ID   Q7PMG3      PRELIMINARY;      PRT;   467 AA.
AC   Q7PMG3;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000011853.
GN   Name=ENSANGG00000009364;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008980; EAA13882.2; -.
SQ   SEQUENCE   467 AA;  54183 MW;  20285C3C254EEF6A CRC64;

  Query Match             17.3%;  Score 458;  DB 2;  Length 467;
  Best Local Similarity   30.0%;  Pred. No. 7.5e-26;
  Matches  121;  Conservative   82;  Mismatches  185;  Indels   16;  Gaps    9;

Qy         65 RLDADYEAPP-LDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGPA-- 121
              :|: | |:|  : |:|   ::  |  :   | ||     :    :|:||:||:|
Db         38 QLEQDNESPTHMYRRKLKIASDVNLLDQHDS-FYHTTKSLLVLFQIMGVMPIMRSPKGVD 96

Qy        122 --RAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGP-FEEAVIAYLFLVNIL 178
                |  |   | :|:::   :       |  ||  ||:  | |   |:| :  :|:  ::
Db         97 MPRTTFTWCSKAFLWAYFIYACETVIVLVVARERINKFISTSDKRFDEVIYNIIFMSIMV 156

Qy        179 PIMIIPIL-WYEARKIAKLFNDWDDFEVLYYQISGHSLPL-KLRQKAVYIAIVLPILSVL 236
              |  ::|:  |    ::||  | | ||:  |  ::|  :   ||     : ||   ||::
Db        157 PHFLLPVASWRNGSEVAKFKNMWTDFQYKYLIVTGKPIVFPKLYPITWTLCIVSWSLSLV 216

Qy        237 SVVITHVTMSDLNINQVVPYCILDNLTAMLGAW---WFLICEAMSITAHLLAERFQKALK 293
               ::  :    |         |    :: |||  :   ||: | |    :   |:     |
Db        217 IILSQYYLQPDFQFCHTFAYY---HIIAMLNGFCSLWFVNCTAFGTASKAFAKELTDVLA 273

Qy        294 HIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSE-GF 352
               ||| : :|| ||: || : :  | |    :     | :||   :: || :|:: | |
Db        274 TERPAAKLTEYRHLWVDLSHMMQQLGKAYSNMYGIYCLVIFPTTIIATYGSLSEIIEHGA 333

Qy        353 GIKDIGLTITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFL 412
               |::|| :   : : ||| ||:|||:||  |  |||::|| | |  ::    | |: |||
Db        334 TYKEVGLFVIVFYCMSLLFIICNEAHHASKRVGLNFQERLLNVNLTAVDKATQKEVEMFL 393

Qy        413 RATEMNPSTINCGGPFFDVNRTLFKGLLTTMVTYLVVLLQFQISI 456
              | : || |:|  |: ::|| |    :: | ||||||:||::::
Db        394 VAIDKNPPTMNLDGYANINRGLITSNISFMATYLVVLMQFKLTL 437


RESULT 5
G21A_DROME
ID   G21A_DROME      STANDARD;      PRT;   447 AA.
AC   Q9VPT1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   25-JAN-2005 (Rel. 46, Last annotation update)
DE   Putative gustatory receptor 21a.
GN   Name=Gr21a; Synonyms=GR21D.1; ORFNames=CG13948;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
```

```
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   GENOME REANNOTATION.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [3]
RP   IDENTIFICATION, AND TISSUE SPECIFICITY.
RX   MEDLINE=20175760; PubMed=10710312; DOI=10.1126/science.287.5459.1830;
RA   Clyne P.J., Warr C.G., Carlson J.R.;
RT   "Candidate taste receptors in Drosophila.";
RL   Science 287:1830-1834(2000).
RN   [4]
RP   IDENTIFICATION.
RX   MEDLINE=21407712; PubMed=11516643; DOI=10.1016/S0960-9822(01)00258-5;
RA   Dunipace L., Meister S., McNealy C., Amrein H.;
RT   "Spatially restricted expression of candidate taste receptors in the
RT   Drosophila gustatory system.";
RL   Curr. Biol. 11:822-835(2001).
CC   -!- FUNCTION: Probable role in the gustatory response.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- TISSUE SPECIFICITY: Expressed in the adult labellar chemosensory
CC       neurons.
CC   -!- SIMILARITY: Belongs to the G-protein coupled receptor Dr-tr
CC       family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE003588; AAF51461.2; -.
DR   FlyBase; FBgn0041250; Gr21a.
DR   GO; GO:0016021; C:integral to membrane; NAS.
DR   GO; GO:0008527; F:taste receptor activity; NAS.
DR   GO; GO:0050909; P:perception of taste; NAS.
KW   G-protein coupled receptor; Glycoprotein; Hypothetical protein;
KW   Multigene family; Receptor; Transmembrane.
FT   DOMAIN        1    107       Extracellular (Potential).
FT   TRANSMEM    108    128       1 (Potential).
FT   DOMAIN      129    146       Cytoplasmic (Potential).
FT   TRANSMEM    147    167       2 (Potential).
FT   DOMAIN      168    199       Extracellular (Potential).
FT   TRANSMEM    200    220       3 (Potential).
FT   DOMAIN      221    230       Cytoplasmic (Potential).
FT   TRANSMEM    231    251       4 (Potential).
FT   DOMAIN      252    305       Extracellular (Potential).
FT   TRANSMEM    306    326       5 (Potential).
FT   DOMAIN      327    338       Cytoplasmic (Potential).
FT   TRANSMEM    339    359       6 (Potential).
FT   DOMAIN      360    415       Extracellular (Potential).
FT   TRANSMEM    416    436       7 (Potential).
FT   DOMAIN      437    447       Cytoplasmic (Potential).
FT   CARBOHYD    378    378       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   447 AA;  51907 MW;  D8B3356422E731DE CRC64;

  Query Match             15.3%;  Score 405;  DB 1;  Length 447;
  Best Local Similarity   26.1%;  Pred. No. 6.4e-22;
  Matches  115;  Conservative   89;  Mismatches  190;  Indels   46;  Gaps   11;

Qy         42 PLNSANAQAYLYGVRKY--SIGLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRN 99
              |: : | : :|    :|   : |  | ||  |      ::   |    :    |   ||:
Db         18 PMLNPNQRQFLEDEVRYREKLKLMARGDAMEEVYVRKQETVDDPLELD----KHDSFYQT 73

Qy        100 IDPINWFLRIIGVLPIVRHGP----ARAKFEMNSASFIYSVVFFVLLACYVGYVANNRI- 154
                :    :|:||:|| |: |      |  :   |   ::::  :      |  |   |:
Db         74 TKSLLVLFQIMGVMPIHRNPPEKNLPRTGYSWGSKQVMWAIFIYSCQTTIVVLVLRERVK 133

Qy        155 HIVRSLSGPFEEAVIAYLFLVNILPIMIIPIL-WYEARKIAKLFNDWDDFEVLYYQISGH 213
               | |    |:||:  :|:  :    ::|:  |    ::|   | | :::  ::: :|
Db        134 KFVTSPDKRFDEAIYNVIFISLLFTNFLLPVASWRHGPQVAIFKNMWTNYQYKFFKTTGS 193

Qy        214 SLPLKLRQKAVYIAIVLPILSVLS---VVITHVTMSDLNINQVVPYCILDN--------- 261
                           || | |  |:     | : :    :|::|   | :  :
Db        194 -------------PIVFPNLYPLTWSLCVFSWLLSIAINLSQ---YFLQPDFRLWYTFAY 237

Qy        262 --LTAMLGAW---WFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTR 316
                 : |||  :   |:: | |    :  |::  |  ::   ||  : :|| ||: || : :
Db        238 YPIIAMLNCFCSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQ 297

Qy        317 DTGNALCYTFVFMSLYLFFIITLSIYGLMSQ-LSEGFGIKDIGLTITALWNIGLLFYICD 375
               | |    :    | :||   :: || :|: :  |   |::|| :   : :|||: ||:
Db        298 QLGRAYSNMYGMYCLVIFFTTIIATYGSISEIIDHGATYKEVGLFVIVFYCMGLLYIICN 357

Qy        376 EAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLF 435
              |||||  |  :|| ||| : |  :::  | |: | | |   ||  :|  |: ::|| |
Db        358 EAHYASRKVGLDFQTKLLNINLTAVDAATQKEVEMLLVAINKNPPIMNLDGYANINRELI 417

Qy        436 KGLLTTMVTYLVVLLQFQIS 455
                :: | |||||||||:|:
Db        418 TTNISFMATYLVVLLQFKIT 437


RESULT 6
G98D_DROME
ID   G98D_DROME     STANDARD;      PRT;   412 AA.
AC   Q8IMN5;
DT   05-JUL-2004 (Rel. 44, Created)
```

```
DT   05-JUL-2004 (Rel. 44, Last sequence update)
DT   25-JAN-2005 (Rel. 46, Last annotation update)
DE   Putative gustatory receptor 98d.
GN   Name=Gr98d; Synonyms=GR98B.1; ORFNames=CG31061;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   GENOME REANNOTATION.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [3]
RP   IDENTIFICATION.
RX   MEDLINE=20175760; PubMed=10710312; DOI=10.1126/science.287.5459.1830;
RA   Clyne P.J., Warr C.G., Carlson J.R.;
RT   "Candidate taste receptors in Drosophila.";
RL   Science 287:1830-1834(2000).
CC   -!- FUNCTION: Probable role in the gustatory response.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: Belongs to the G-protein coupled receptor Dr-tr
CC       family. Subfamily I.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE003761; AAN14112.1; -.
DR   FlyBase; FBgn0046885; Gr98d.
KW   G-protein coupled receptor; Glycoprotein; Hypothetical protein;
KW   Multigene family; Receptor; Transmembrane.
FT   DOMAIN        1      7       Extracellular (Potential).
FT   TRANSMEM      8     28       1 (Potential).
FT   DOMAIN       29     40       Cytoplasmic (Potential).
FT   TRANSMEM     41     61       2 (Potential).
FT   DOMAIN       62    148       Extracellular (Potential).
FT   TRANSMEM    149    169       3 (Potential).
FT   DOMAIN      170    181       Cytoplasmic (Potential).
FT   TRANSMEM    182    202       4 (Potential).
FT   DOMAIN      203    244       Extracellular (Potential).
FT   TRANSMEM    245    265       5 (Potential).
FT   DOMAIN      266    294       Cytoplasmic (Potential).
FT   TRANSMEM    295    315       6 (Potential).
FT   DOMAIN      316    369       Extracellular (Potential).
FT   TRANSMEM    370    390       7 (Potential).
FT   DOMAIN      391    412       Cytoplasmic (Potential).
FT   CARBOHYD      4      4       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    127    127       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    222    222       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   412 AA;  47722 MW;  7B05D0AF4FAC4D72 CRC64;

  Query Match             6.3%;  Score 167.5;  DB 1;  Length 412;
  Best Local Similarity   21.2%;  Pred. No. 0.00031;
  Matches   88;  Conservative   74;  Mismatches  140;  Indels  113;  Gaps   17;

Qy        102 PINWFLRIIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNR-IHIVRS- 159
              || :| | :       |   |    :  | :: |:   |:  ||   ||  : ||   :
Db         27 PIQFFTRTL-------HKRRRGIVILGYACYLISISLMVIYECYANIVALQKDIHKFHAE 79

Qy        160 ----LSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEV-LYYQISG-- 212
                 :|  :: ::  :|: | | |::         |::|::::|  | |: |    ||
Db         80 DSSKVMGNTQKVLVVAMFVWNQLNILL------NFRRLARIYDDIADLEIDLNNASSGFV 133

Qy        213 ---HSLPLKLR-QKAVYIAIVL-------------PILSVLSVVITHVTMSDLNINQVV 254
                 |   : |  :| : |||               | |   : |:| : :  | : :
Db        134 GQRHWWRFRFRLALSVGLWIVLLVGLTPRFTLVALGPYLHWTNKVLTEIILIMLQL-KCT 192

Qy        255 PYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKL 314
              ||:          :  || | :    |:| ::    |:
Db        193 EYCV----------FVLLIYELILRGRHIL-QQISVELE-----------------GNQ 223

Qy        315 TRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYIC 374
              :||:   ||     ::|    ::   |:||::::|  |    : ||:  |:|   :  |
Db        224 SRDSVQELC-----VALKRNQLLAGRIWGLVNEVSLYF---TLSLTLLFLYNELTILQIV 275

Qy        375 DEAHYASVNVRTNFQKK----LLMVELNWMNSDAQTE---------------------- 407
              : |  |||     | :     |:: :|    |   :|
Db        276 NWALIKSVNPNECCQYRRVGTCLLLSINIFLSCLYSEFCIQTYNSISRVLHQMYCLSAAE 335

Qy        408 ----INMFLRATEMNPSTI----NCGGFPDVNRTLFKGLLTTMVTYLVVLLQFQI 454
                  : | ||   :    :     ||| ||:|   | |:: |:  |:::|:||:|
Db        336 DYLILKMGLREYSLQMEHLKLIFTCGGLFDINLKPFGGMVVTLPGYIIILVQFKI 390


RESULT 7
Q7PVK1
ID   Q7PVK1      PRELIMINARY;      PRT;   383 AA.
AC   Q7PVK1;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
```

```
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000024184.
GN   Name=ENSANGG00000020468;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008984; EAA43431.1; -.
SQ   SEQUENCE   383 AA;  44507 MW;  703E090992D3B5B4 CRC64;

  Query Match             5.7%;  Score 149.5;  DB 2;  Length 383;
  Best Local Similarity   20.4%;  Pred. No. 0.0063;
  Matches   81;  Conservative   78;  Mismatches  156;  Indels   83;  Gaps   18;

Qy        109 IIGVLPIVRHGPARAKFEMNSASFIY--SVV-------FFVLLACYVGYVANNRIHIVRS 159
              ::||:|   |  :    |:     :|  ::|       |:: |  ::      |:|
Db         17 LLGVVPFHFH-RSTLTLELRRELLLYCCAIVILGPTGRFYIYLTTFL------RMH---- 65

Qy        160 LSGPFEEAVIAYLFL-VNILPIMIIP--ILWYEARKIAKL-----------------FND 199
              | |  |  : ||: :: : ::|:|   | |  ::: :|                    |
Db         66 --GKFIVAFYSALFIELSDVVLLIVPPCCLLYRHKQVKRLFVLLVKLYHNPAVSAFPFRR 123

Qy        200 WDD----FEVLYYQISGHSLPLKLRQ--KAVYIAIVLPILSVLSVVITHVTMSDLNINQV 253
              |      | ::|    | :  | : |     :: ||  :||          :| |: : |
Db        124 WFRTCLLFNLMYESAYGVNFLLNIFQPGSTFHLEIVNLLLS--------TALSQLSKSAV 175

Qy        254 VPYCILDNLTAMLGAWWFLICEAMSITAHLL-------AERFQKALK--HIGPAAMVADY 304
                  :  | |:  :|    : :  |||           |: |:  |: ||  :  |
Db        176 --------MLIMYGSGQYLQMALVQLKEHLLLGVHRSTGPVQQETLRQLHVRPAGWMQYY 227

Qy        305 RVLWLRLSKLTRDTGNALCYTFVFMSL-YLFFIITLSIYGLMSQLSEGFGIKDIGLTITA 363
                 || :: ::  |        |:  ||| :| ||  : :  :| :      ::::   :
Db        228 GQLW-QVCEILNDLFGVPLITYCLMSLMHLTFICYVMLTKMMVEWHNFSVVQNVLAGVEL 286

Qy        364 LWNIGLLFYICDEAHYASVNVRT-----NFQKKLLMVELNWMNSDAQTEINMFLRATEMN 418
              :| :   | :        |   ||        || || : |: |:   :  | :|   |
Db        287 MWTV---FELLCLMRIVSTFARTREESLKTQKLLLRLNLSPMDHKLKQSIEVFALQTLHQ 343

Qy        419 PSTINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISI 456
              |        | :: |:   :   :  ||::|:||:::|
Db        344 PIEFTACRMFTLDYTVLFSIAAAVTNYLIILIQFEMAI 381


RESULT 8
Q7PEV2
ID   Q7PEV2      PRELIMINARY;      PRT;   472 AA.
AC   Q7PEV2;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000023700.
GN   Name=ENSANGG00000003253;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008794; EAA45578.1; -.
DR   InterPro; IPR009318; Trehalose_recept.
DR   Pfam; PF06151; Trehalose_recp; 1.
SQ   SEQUENCE   472 AA;  54427 MW;  EDA32CC87EBE4388 CRC64;

  Query Match             5.6%;  Score 148.5;  DB 2;  Length 472;
  Best Local Similarity   23.0%;  Pred. No. 0.0095;
  Matches  102;  Conservative   62;  Mismatches  162;  Indels  117;  Gaps   20;

Qy         96 FYRNIDPINWFLRIIGVLPI---VRHGPARAKFEMNSASFI--YSVVFFVLLACYVGYVA 150
              | | | |     :: |: ||    |: | | : :   |   |   :||   ||  | |
Db         64 FLRGIRPFVILGQMFGIFPIYGVTRNDPKRFRLKWFSLRVILNLTVVVTALLQAYYEYGR 123

Qy        151 NNRIHI----VRSLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEVL 206
              | |    | ||    :  :|  ||| |:            | |   :   ||: :
Db        124 LKAIGINAKNVSSLIFFIDACLINVLFL-NL-----------ATKWRSVAMKWDEVDDT 170

Qy        207 YYQISGHSLPLKLRQK-------AVYIAIVLPILSVLSVVITHVTMSDL----------- 248
              : :   |     ||::        |::| |  |||::| |  |  | ::
Db        171 FNRPPYHMQSWSLRKRLGVVSFTLVFLAAVEHILSIVSNV--HNQMVEIKYCNWTEPNYF 228

Qy        249 ---------NINQVVPYCILD-----NLTAMLGAWW-----FLICEAMSITAHLLAERFQ 289
                    . ||    ||  |       ::: |  :|     |:|  ::      || |||
Db        229 QHYSLRRFANIYLNFPYNSLSAVPFTYVSSALTMYWNYQDIFIIMISIG-----LATRFQ 283

Qy        290 KA---LKHIGPAAMV--ADYRVLWLR-------LSKLTRDTGNALCYTFVFMSLYLFFII 337
              :    || :    ::    |:   |:|       : :|  |   |: :| :       : |
Db        284 QINNYLKILSDGVLIPGEDF---WIRVRTNYVAVCELLDDVDRAISWTMLISCATNLYYI 340

Qy        338 TLSIYGLMSQLSEGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVRT----------- 386
              | |  :  :|:    ::|       | :   | | |           |||
Db        341 CLQILHVSKKLAN--TVED------AYYGFSLGFLI----------VRTVIVFLSAAHIH 382

Qy        387 NFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYL 446
              : || | : :   |       |:  |   :     ::  ||| : | |   :  |:|||
Db        383 DCAKKPLDIIMKIPNVGWCVELERFSTQLKSEKVALSGMGFFSLTRQLLFSMAGTIVTYE 442

Qy        447 VVLLQFQISIPTDKGDSEGANNI 469
              :|:|:|         :||   ||
Db        443 LVMLKFD-------QESESKGNI 458


RESULT 9
Q7PVJ9
ID   Q7PVJ9      PRELIMINARY;      PRT;   404 AA.
AC   Q7PVJ9;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000024208.
GN   Name=ENSANGG00000020468;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008984; EAA43429.1; -.
DR   InterPro; IPR000169; Pept_cys_acsite.
DR   PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
SQ   SEQUENCE   404 AA;  46605 MW;  CB0C0C4F7EEC5E63 CRC64;

  Query Match             5.5%;  Score 145.5;  DB 2;  Length 404;
  Best Local Similarity   20.3%;  Pred. No. 0.013;
  Matches   78;  Conservative   84;  Mismatches  165;  Indels   57;  Gaps   14;

Qy        110 IGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAVI 169
              :||:|   | ||       |  |  |:: :  :   :| :    :::    :  |  :|
```

```
Db         39 LGVVPYEFH-PATL-----SLQFRRSLLLYCFIIAIIGPMGRCWVYLTTFFTMNFGVSVA 92

Qy        170 AY----LFLVNILPIMIIP-ILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAV 224

Db         93 HYTAMFMELLDAILLVVPPCYLLCRHKKVKELFVLL--VKVYYSSLSLSIFPLRLWFRAC 150

Qy        225 YIAIVLPILSVLSVVITHVTMSDLNINQVVP----------YCILDNLTAML----- 266

Db        151 FL---------FNLVYDSAYGVNVALNLVLPGTGIHEKYVKILSWSVLSQLTKSIVLLII 201

Qy        267 -GAWWFLICEAMSITAHLLAER------FQKALK--HIGPAAMVADYRVLWLRLSKLTRD 317

Db        202 YGTAQYLLRMVDHLNDHLAAGRLSTNTVYQEQIRQQHHGPAGYLEFYEQLG-KMSETLND 260

Qy        318 TGNALCYTFVFMSL-YLFFIITLSIYGLMSQLS----EGFGIKDIGLTITALWNIGLLFY 372

Db        261 LYGVPLITYFLMTLIHLTFVCYMTLTKVLTKWSYLSWHALLIAGIAVTSNVIDLLCLMKI 320

Qy        373 ICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPSTINCGGFFDVNR 432

Db        321 VGTFARTREESLKT--QKLLLRLNLSPMDHKLKQSIEVFALQTLHQPIEFTACRMFTLDY 378

Qy        433 TLFKGLLTTMVTYLVVLLQFQISI 456

Db        379 TVLFSIAAAVTNYLIILIQFEMAI 402


RESULT 10
Q7PPB5
ID   Q7PPB5      PRELIMINARY;      PRT;   444 AA.
AC   Q7PPB5;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000004263.
GN   Name=ENSANGG00000003364;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008960; EAA10739.2; -.
DR   InterPro; IPR009318; Trehalose_recept.
DR   Pfam; PF06151; Trehalose_recp; 1.
SQ   SEQUENCE   444 AA;  51486 MW;  6FE09ACD0EE059A9 CRC64;

  Query Match             5.4%;  Score 143;  DB 2;  Length 444;
  Best Local Similarity   21.5%;  Pred. No. 0.023;
  Matches   98;  Conservative   73;  Mismatches  170;  Indels  114;  Gaps   24;

Qy         59 SIGLAERLDADYEAPPLD---RKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPI 115

Db         26 STALAE------ETPPADGAAERECSTHEAVAAVIFMGQLF--SLIPIDGYAR------- 70

Qy        116 VRHGPARAKFEMNSASFIYS-VVFFVLLACYVGYVANNRIHIVRSLSGPFEEA--VIAYL 172

Db         71 -STDPRDVRMRLRSVQFVYGCVTLFIMLTLIIMLC----VHTAHEPSFGVQQATSLVYYA 125

Qy        173 FLV-NILPIMIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLP 231

Db        126 IIVFFMVELMLLARNW--SQIMGRWYTDEAPFRTDPYRPPSRTLP--FRRKVHLIAFGVM 181

Qy        232 ILSVLSVVITHVTMSDLN----------------------INQVVPYCILDNLTAMLG 267

Db        182 FLAFVEDTLNFVSAYRLNELHIRYCPHTAGFWKNFFHREHPYVLRVIPYHPVVGWTIELT 241

Qy        268 ------AWW----FLICEAMSITAHLLAERFQKALKHI-GPAAMVADYRVL---WLRLSK 313

Db        242 MRIAKFTWHYVDVFIICLSLGLQRRFV--QFNERLERLDGQPQSQGVWRALRLDYVRLSE 299

Qy        314 LTRDTGNALCYTFV---FMSLYLF------FIITLSIYGLMSQLSEGFGIKDIGLTITAL 364

Db        300 LV---------TFVDERFSKLILFCCANDMFFITV-------QLFNSFDLKPTTVTTVYF 343

Qy        365 W-NIGLLFYICDEAHYASVNVRTNFQKKLLMV----ELNWMNSDAQTEINMFLR----AT 415

Db        344 WYSLGFLIGRCFLMLFVVSSISRASEKPLETLRRFPSTNW---------NLDLRRLCDAV 394

Qy        416 EMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLL 450

Db        395 ATSENALSGKRFFFVRRPLILAMAGTIITYELVLL 429


RESULT 11
Q7PKI4
ID   Q7PKI4      PRELIMINARY;      PRT;   416 AA.
AC   Q7PKI4;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000023329.
GN   Name=ENSANGG00000020045;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008987; EAA43247.1; -.
SQ   SEQUENCE   416 AA;  48071 MW;  1A6030890A0F68B1 CRC64;

  Query Match             5.4%;  Score 142;  DB 2;  Length 416;
  Best Local Similarity   19.6%;  Pred. No. 0.025;
  Matches   81;  Conservative   80;  Mismatches  182;  Indels   70;  Gaps   15;

Qy        100 IDPINWFLR----------------------IIGVLPIV--RHGPARAKFEMNSASFIY 134

Db         12 VSPVRWSVRRFSCCTMQVPRWLIVLQSVVLKFLFAILPLSFDRH---RERFRNRTKDVIL 68

Qy        135 SVVPFVLLACYVGYVANNRIHIVR-SLSGPFEEAVIAYLFLVNILPIMIIPILWYEARK- 192

Db         69 STVLLVLYA-IVGPIVCVELYVMNISVENQITNALSAFQFAFVYFFILIVQIIPIRKKAC 127

Qy        193 IAKLFNDWDDFEVLYYQISGHSLPLKLRQKAVYIAIVLPILSVLSVVITHVTMSDLNINQ 252

Db        128 LHVLLNEMFQLKHVLERMVARSM-LEFRLYRLFLVKVTAV-DILMQLLALYTFEDSTYTG 185

Qy        253 VVPYCILDNLTAMLGAWWFLICEAMSITAHLLAERFQKALKHIGPAAM------------ 300

Db        186 RVDSVGLASWLTFYAMMYFIAVIENFILAGVLIGAVMQAMVNIGLKRLARSRQGSAAGRQ 245

Qy        301 --VADYRVLWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLSEGFGIKDIG 358

Db        246 PSVLQLYMLHCKNEDMVKQFMETLNFPTLMLTGWYFFMIVYSVRPPASPV--GNGIAD-- 301

Qy        359 LTITALWNIGLLF-YICDEAHYASV--NVRTNFQKKLLMVELNWMNSDAQTE-------- 407

Db        302 -DFKAYINPLIFFLYQCVQLYLLVLIPSVYTDHAKKMMRL-LNYVSVNQHHHRPGQERLV 359

Qy        408 ------INMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQI 454

Db        360 GREGGVIEVLMVDCMQRNNAINNYGMYAMNRALLFGMIATMTSYLIILIQFHI 412
```

```
RESULT 12
Q7PVK0
ID   Q7PVK0      PRELIMINARY;      PRT;   411 AA.
AC   Q7PVK0;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000023884.
GN   Name=ENSANGG00000020468;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008984; EAA43430.1; -.
SQ   SEQUENCE   411 AA;  47111 MW;  6D7A379CB2729E5E CRC64;

  Query Match             5.3%;  Score 141.5;  DB 2;  Length 411;
  Best Local Similarity   21.8%;  Pred. No. 0.027;
  Matches   87;  Conservative   64;  Mismatches  161;  Indels   87;  Gaps   16;

Qy        110 IGVLPIVRHGPARAKFEMNSASFIYSVVFFV---LLACYV----GYVANNRIHIVRSLSG 162
              :||:|    | ||    : : :  :|  :  :   |  | |     :     | :||  :
Db         46 LGVVPYEFH-PATLSLQFSRSLLVYCFIIAIISPLGRCCVFLTTFFTMKLGISVVR-YTA 103

Qy        163 PFEEAVIAYLFLVNILPIMIIP--ILWYEARKIAKLFNDWDDFEVLYYQISGHSLPL--- 217
              | | : | |        ::||   |  :  |: :||        |  :|   ||||
Db        104 MFRELLDAIL--------LVIPPCYLLCQHAKVKQLFE-------LSVEIYRSSLPLGIF 148

Qy        218 ------------KLRQKAVY-----IAIVLPILSVLSVVITHVTMSDLNINQVVPYCILD 260
                         |   :||       :|||   : :  :            |:: : :|
Db        149 PFWQWFRTCLLCSLVYDSVYGVNVAFNLVLPGAGIHAKYV-----------QILLWSVLS 197

Qy        261 NLT------AMLGAWWFLICEAMSITAHLLAER------FQKALK--HIGPAAMVADYRV 306
               ||       : |   :|:      :  ||:| |       :|: ::  | |||  :  |
Db        198 QLTKSVVLLIIYGTAQYLLMMVDQLKDHLVAGRLSTNPVYQEQIRQQHHGPAGYLMFYEQ 257

Qy        307 LWLRLSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLM---------SQLSEGFGIKDI 357
              |      |   |  |   |:   ::| |:  | :  |:          | |  |  |
Db        258 LAKVCETLNDLIGVPLITYFLMALIHLTFVCYLILTKLLVRWSYITWQSILIAGIAIASY 317

Qy        358 GLTITALWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEM 417
               : :  |  |   |      |      :::|  || || : |: |:   :  | :|   |
Db        318 VIDLLCLMRIVGTF-----ARTREESLKT--QKLLLRLNLSPMDHKLKQSIEVFALQTLH 370

Qy        418 NPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISI 456
                |        | :: |:   :   :  ||::|:||:::|
Db        371 QPIEFTACRMFTLDYTVLFSIAAAVTNYLIILIQFEMAI 409


RESULT 13
Q7PVJ8
ID   Q7PVJ8      PRELIMINARY;      PRT;   382 AA.
AC   Q7PVJ8;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000024480.
GN   Name=ENSANGG00000020468;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008984; EAA43433.1; -.
SQ   SEQUENCE   382 AA;  43445 MW;  8863D72B4E834FBC CRC64;

  Query Match             5.3%;  Score 140;  DB 2;  Length 382;
  Best Local Similarity   20.4%;  Pred. No. 0.032;
  Matches   81;  Conservative   78;  Mismatches  155;  Indels   84;  Gaps   18;

Qy        109 IIGVLPIVRHGPARAKFEMNSASFIYSVVFFVLLACYVGYVANNRIHIVRSLSGPFEEAV 168
              ::||:|:      |: :|:            :|  ||   :|   :| :  |:  |   :
Db         17 LVGVIPL---RLPRSSWEL----------LILCYCYTIAIAGPLVHFLLYLAW-FVIVM 61

Qy        169 IAY--LFLVNI------LPIMIIPILWY-----EARKIAKLFNDWDDFEVLYYQISGHS- 214
              : :   |:|:|      :| :| |  :      : :::|:|       :|    | |:
Db         62 VKFRVTFVVSICMHLGEVPFLIFPPCYLLSCRAKMKQLAELCTHVHSSSLLVGPFSLHAW 121

Qy        215 ----LPLKLRQKAVYIAIVLPILSVLSV----------VITHVTMSDLNINQVVPYCIL 259
                      |  :| :   :|  |:: ||           :  : :|:|  : |:
Db        122 YRKCFVFNLGYEAAHTVTLLANLALQSVFPEGMGKIYPQTLLWIALSNLTKSAVL----- 176

Qy        260 DNLTAMLGAWWFLICEAMSITAHLLAERF--------QKALK--HIGPAAMVADYRVLWL 309
                ||      : :||   ::  : |  |          |: |:  | |   ::  |  ||
Db        177 --LTMYGSVQYLMIC---AVKLNELLPRLSSWTDPYRQETLRRLHFGRDGLMDFYGQLW- 230

Qy        310 RLSKLTRDTGNALCYTFVFMSLYLFFIITLSI-YGLMSQLSEGFG----IKDIGLTITAL 364
                    :  | |  :   | : || | :| | : |||: |       :  : | :  :
Db        231 ----QGAELANELFGAPLLAYLAIAFIHTTTIYYRVWSQLAGGLENQPLVITVHLIVQLV 286

Qy        365 W-NIGLLFYI-----CDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMN 418
              |  : |:| :     | :    |:      || || : |: |:   :  | :|   |
Db        287 WVTMDLMFLMAVVGSCGQMRQESLKT----QKLLLRLNLSPMDHKLKQSIEVFALQTLHQ 342

Qy        419 PSTINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISI 456
              |        | :: |:   :   :  ||::|:||:::|
Db        343 PIEFTACRMFTLDYTVLFSIAAAVTNYLIILIQFEMAI 380


RESULT 14
Q7PKC5
ID   Q7PKC5      PRELIMINARY;      PRT;   451 AA.
AC   Q7PKC5;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   ENSANGP00000022858.
GN   Name=ENSANGG00000021334;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
DR   EMBL; AAAB01008986; EAA43315.1; -.
SQ   SEQUENCE   451 AA;  51724 MW;  4CF7E1BBDBB49A84 CRC64;

  Query Match             5.3%;  Score 139.5;  DB 2;  Length 451;
  Best Local Similarity   18.7%;  Pred. No. 0.042;
  Matches   86;  Conservative   95;  Mismatches  174;  Indels  105;  Gaps   21;

Qy         72 APPLDRKKSSDSTASNNPEFKPSVFYRNIDPINWFLRIIGVLPIVRHGPARAKFEMNSAS 131
              |||      :  :|      ||:  :|:         |     |::        : :: : :
Db         16 APPSKPGPALQATPVGEQEFE-QLFHFAFKCFRLFALTPGLM-----DRQKDRYRVRNTR 69
```

```
Qy        132 FIYSVVFFVLLACYVG----YVANNRIHIVRSLSGPFEEAVIAYLFLVNILPI---MIIP 184
Db         70 WMMLIVLLLVVVAWIALFETFPFIERRTALITGIANHIQ-------FLMNTIALTVAWVVP 122

Qy        185 ILWYEARKIAKLFNDWDDFEVLYYQISGHS---------LPLKLRQKAVYIAIVLPILSV 235
Db        123 QL-----KADELGSILDGFLLIDRELSSYNVHEVAGYKRVSFLLRYGVVLLALM--SLTV 175

Qy        236 LSVVITHVTMSDLNI----NQVVPYCI--LDNLTAMLGAWWF--------LICEAMSITA 281
Db        176 YDGFVSFVQLTTVEVWYWLSHQLPFIIYAMAFLHAYVLIYWLHARFRRLNTLVEQYYRQG 235

Qy        282 HLLAERFQ-----------------KALKHIGPAAMVADYRVLWLRLSKLTRDTGNALC 323
Db        236 HIFAPARQTIISFATMVKLDEESAVGEELHSVGRREISDDLQV--LAIVSRTIDLGQKIE 293

Qy        324 YTF--VFMSLY--LFFIITLSIY----GLMSQLSEGFGIKDI----GLTITALWNIGLLF 371
Db        294 SYFGPLFLTVYTALFSVTTIQSYYCYLHLTAKGDRGLSIETLVLSGGIILYNVIAIVALP 353

Qy        372 YICDEAHYASVNVRTNFQKKLLM---VELNWMNSDAQTEINMFLRATEMNPSTI-----N 423
Db        354 YICEQVESES---------KLLMSYLSKLSMKHSQVAQHSSIWF------PNLISSVRFS 398

Qy        424 CGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPTDKGDS 463.
Db        399 AFGFFTINYNMLSGLVAGMVTYLIIFIQFNSMVPAGKDDT 438
```

RESULT 15
Q7QKL0
ID   Q7QKL0   PRELIMINARY;   PRT; 1210 AA.
AC   Q7QKL0;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   EbiP4099 (Fragment).
GN   Name=ebiG4099; ORFNames=ENSANGG00000003253;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC   EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC   preliminary data.
DR   EMBL; AAAB01008794; EAA03574.1; -.
DR   InterPro; IPR009318; Trehalose_recept.
DR   Pfam; PF06151; Trehalose_recp; 2.
FT   NON_TER   1   1
FT   NON_TER   1210   1210
SQ   SEQUENCE   1210 AA; 140580 MW; 6DF7CA521861C008 CRC64;

  Query Match   5.3%; Score 139.5; DB 2; Length 1210;
  Best Local Similarity   22.7%; Pred. No. 0.13;
  Matches   98; Conservative   62; Mismatches 154; Indels 117; Gaps   20;

```
Qy        108 RIIGVLPI---VRHGPARAKFEMNSASFI--YSVVFFVLLACYVGYVANNRIHI----VR 158
Db         12 QMFGIFPIYGVTRNDPKRFRLKWFSLRVILNLTVVVTALLQAYYEYGRLKAIGINAKNVS 71

Qy        159 SLSGPFEEAVIAYLFLVNILPIMIIPILWYEARKIAKLFNDWDDFEVLYYQISGHSLPLK 218
Db         72 SLIFFIDACLINVLFL-NL------------ATKWRSVAMKWDEVDDTFNRPPYHMQSWS 118

Qy        219 LRQK-------AVYIAIVLPILSVLSVVITHVTMSDL-------------------NIN 251
Db        119 LRKRLGVVSFTLVFLAAVEHILSIVSNV--HNQMVEIKYCNWTEPNYFQHYSLRRFANIY 176

Qy        252 QVVPYCILD-----NLTAMLGAWW-----FLICEAMSITAHLLAERFQKA---LKHIGPA 298
Db        177 LNFPYNSLSAVFFTYVSSALTMYWNYQDIFIIMISIG-----LATRFQQINNYLKILSDG 231

Qy        299 AMV--ADYRVLWLR-------LSKLTRDTGNALCYTFVFMSLYLFFIITLSIYGLMSQLS 349
Db        232 VLIPGEDF---WIRVRTNYVAVCELLDDVDRAISWTMLISCATNLYYICLQILHVSKKLA 288

Qy        350 EGFGIKDIGLTITALWNIGLLFYICDEAHYASVNVRT-----------NFQKKLLMVELN 398
Db        289 N--TVED------AYYGFSLGFLI----------VRTVIVFLSAAHIHDCAKKPLDIIMK 330

Qy        399 WMNSDAQTEINMFLRATEMNPSTINCGGFFDVNRTLFKGLLTTMVTYLVVLLQFQISIPT 458
Db        331 IPNVGWCVELERFSTQLKSEKVALSGMGFFSLTRQLLFSMAGTIVTYELVMLKFD----- 385

Qy        459 DKGDSEGANNI 469
Db        386 --QESESKGNI 394
```

Search completed: May 23, 2005, 19:25:11
Job time : 177 secs